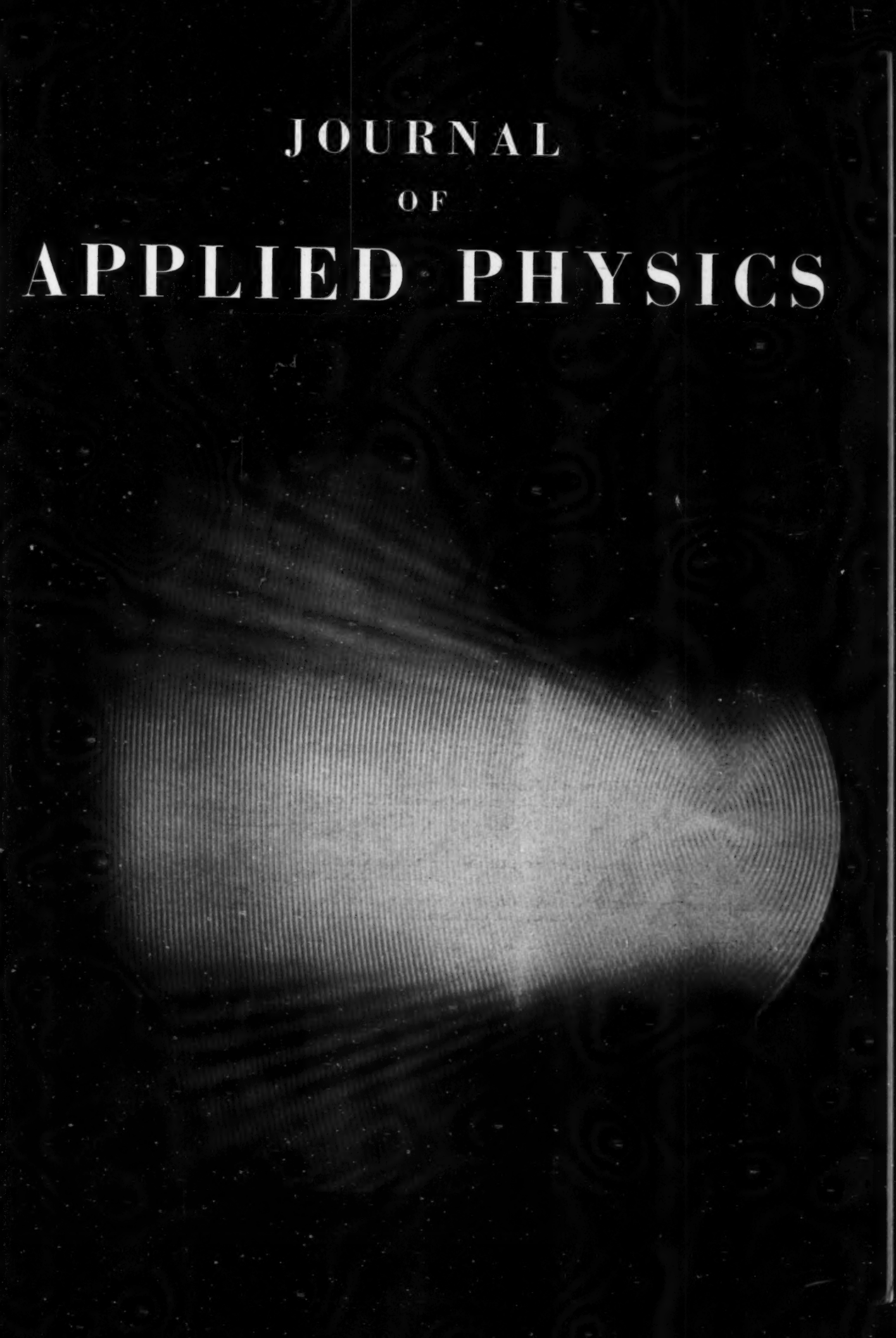
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 20, No. 3 MARCH, 1949

In This Issue

Page

Recent Developments in the Mathematical Theory of Plasticity WILLIAM PRAGER 235

Contributed Original Research

Some Properties of Tubular Electron Beams NELSON WAX 242

New Solution to the Problem of Vertical Angle-of-Arrival of Radio Waves E. W. HAMLIN, P. A. SEAY, AND W. E. GORDON 248

Bending of Beams with Creep E. P. POPOV 251

Analysis of the Metal-Strip Delay Structure for Microwave Lenses S. B. COHN 257

Response of Certain Inorganic Phosphors to Far Ultraviolet Radiation D. H. TOMBOULIAN AND E. M. PELL 263

Phase Lead for A.C. Servo Mechanisms ARTHUR H. BENNER 268

Grain Boundary Model and the Mechanism of Viscous Intercrystalline Slip T'ING-SUI KÊ 274

Disk-Loaded Wave Guides E. L. CHU AND W. W. HANSEN 280

Visual Methods for Studying Ultrasonic Phenomena R. BOWLING BARNES AND CHARLES J. BURTON 286

Letters to the Editor:

Determination of Relative Phases of Fourier Coefficients from X-Ray Diffraction Data DAN MCLACHLAN, JR. 295

Use of Titanium and Zirconium Nitrides as a Conducting Element in Thoria Cathodes O. B. BUSH, R. B. VANDEGRIFT, AND T. E. HANLEY 295

Index to Advertisers xxxiv

BOARD OF EDITORS
ELMER HUTCHISSON, *Editor*

Associate Editors
R. BOWLING BARNES
OTTO BEECK
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
RAYMOND J. SEEGER
FREDERICK SEITZ
CHARLES S. SMITH
J. R. TOWNSEND

Publication of the American Institute of Physics

OFFICERS
GEORGE R. HARRISON, *Chairman*
WALLACE WATERFALL, *Secretary*
GEORGE B. PEGRAM, *Treasurer*
JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD
LUIS W. ALVAREZ
ROBERT B. BRODE
E. U. CONDON
A. J. DEMPSTER
W. F. FAIR, JR.
HARVEY FLETCHER
R. C. GIBBS
GEORGE R. HARRISON
W. V. HOUSTON
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
GEORGE B. PEGRAM
A. H. PFUND (*deceased*)
JOHN T. TATE
WALLACE WATERFALL
M. W. WHITE
A. G. WORTHING

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
CLEVELAND NORCROSS, *General Manager*
THEODORE VORBURGER, *Business Manager*
RUTH F. BRYANS, *Publication Manager*
BETTY HOFSTADTER, *Office Manager*
EILEEN B. NEUBERGER, *Circulation Manager*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $4 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices
Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 20, Number 3 March, 1949

Recent Developments in the Mathematical Theory of Plasticity*

WILLIAM PRAGER
Brown University, Providence, Rhode Island
(Received September 27, 1948)

The first part of the paper (Sections I to V) is concerned with the general technique used in the discussion of stress-strain laws for inviscid elastic-plastic materials with work hardening. It is postulated that when the "mechanical state" of such a material is known a given infinitesimal change of stress produces a uniquely defined infinitesimal change of strain. Once the assumption is made as to which variables determine the mechanical state, the conditions of continuity, uniqueness, irreversibility, and consistency can be used to extract rather far-reaching information regarding the structure of the stress-strain law. The procedure is applied to the case where the mechanical state is determined by the components of stress and permanent strain.

The second part of the paper (Sections VI to VIII) is concerned with various problems of plastic equilibrium. Structural stability in the plastic range is discussed and the difficulties arising in the formulation of stability problems for non-conservative systems are pointed out. Statically determinate problems in the Saint-Venant-Mises theory of plasticity are reviewed with special reference to discontinuous solutions. Finally, the "shake-down" problem is discussed for an elastic-plastic structure under the action of loads, each of which varies in a random manner between given extreme values.

NEXT to dynamics of compressible fluids, the mathematical theory of plasticity is probably the most active field of mechanics of continua at present. A complete account of recent developments in this field would have to deal with two equally important aspects of progress: new solutions to specific problems and new fundamental concepts emerging from the discussion of these problems. Such an account would obviously have exceeded the framework of a General Lecture before this Congress, and a decision had to be made concerning the relative emphasis placed on these two aspects of progress. In stressing the new fundamental concepts in the following survey, the speaker does not wish to belittle the importance of new solutions to specific problems. However, the discussion of most of these problems would require the presentation of a greater amount of mathematical detail than is practical in the course of a single lecture.

I. BASIC CONCEPTS

The strain in a plastic material may be considered as the sum of the reversible *elastic strain* and the permanent *plastic strain*. A material is called *perfectly plastic* or *work-hardening* according to whether, under constant stress, it admits changes of permanent strain or not.

For moderate strains, mild steel behaves approximately like a perfectly plastic material. In view of the importance of mild steel as a structural material, it is not surprising that the classical theory of plasticity is exclusively concerned with perfectly plastic materials. The increasing structural use of aluminum alloys, however, calls for the development of a theory which applies to work-hardening materials. The first step towards such a theory is the establishment of sound stress-strain relations.

To indicate typical arguments used in the discussion of stress-strain laws, let us consider the isothermal deformations of a work-hardening material whose *mechanical state* is defined by the tensors of stress and plastic strain. It is assumed that, when the mechanical state and an infinitesimal change of stress are given, the corresponding infinitesimal change of strain is uniquely determined. Since Hooke's law is assumed to provide the relation between the infinitesimal changes of stress and elastic strain, the stress-strain law for a material of this type reduces, essentially, to a relation involving the instantaneous state and the infinitesimal changes of stress and plastic strain. This relation is assumed to

* This survey which was presented to the Seventh International Congress for Applied Mechanics in London on September 8, 1948, was prepared in the course of research conducted under Contract N7onr-358 sponsored jointly by the Office of Naval Research and the Bureau of Ships, United States Navy.

be homogeneous and linear in the infinitesimal changes of the components of stress and plastic strain. This assumption precludes viscosity effects and thus constitutes an idealization of the fact that, for moderate rates of strain, the mechanical behavior of structural metals is practically independent of the speed with which a test is conducted.

A work-hardening material is said to be in an *elastic state* if any infinitesimal change of stress leaves the permanent strain unchanged; if this condition is not fulfilled, the material is said to be in a *plastic state.*

The *elastic range* associated with a given state is the totality of all elastic states which can be reached from the given state without passing through plastic states. It follows from this definition that the states of an elastic range differ only in stress but have the same permanent strain. The *yield limit* of an elastic range is the totality of the stress tensors of the plastic states which bound the range.

An infinitesimal change of stress from a given plastic state leads either to an elastic state or to another plastic state. In the first case, the change of stress is said to constitute *unloading;* in the second case, it is said to constitute *loading* or to be *neutral* according to whether it is accompanied by a change in permanent strain or not.

II. THE CONDITION OF CONTINUITY

Let σ_x, σ_y, σ_z be the normal stresses and τ_{xy}, τ_{yz}, τ_{zx} the shearing stresses with respect to a set of rectangular axes. Similarly, denote the increments of the plastic extensions and shearing strains by $d\epsilon_x$, $d\epsilon_y$, $d\epsilon_z$ and $d\gamma_{xy}$, $d\gamma_{yz}$, $d\gamma_{zx}$, respectively.

We shall find it useful to discuss stress-strain laws in geometrical terms which refer to a six-dimensional Euclidean space with the rectangular coordinates σ_x, σ_y, σ_z, $\tau_{xy}\sqrt{2}$, $\tau_{yz}\sqrt{2}$, $\tau_{zx}\sqrt{2}$. The stress is thus represented by the radius vector $\boldsymbol{\sigma}$, and an infinitesimal change of stress by an increment $d\boldsymbol{\sigma}$ of this radius vector. Finally, let $d\boldsymbol{\epsilon}$ be the infinitesimal vector with the components $d\epsilon_x$, $d\epsilon_y$, $d\epsilon_z$, $d\gamma_{xy}/\sqrt{2}$, $d\gamma_{yz}/\sqrt{2}$, $d\gamma_{zx}/\sqrt{2}$. The reason for this particular way of representing the stress and the increments of stress and plastic strain will be discussed later.

Let us now consider a given state of a work-hardening material. The yield limit associated with this state is represented by a certain hypersurface in our six-dimensional space. In the following, this hypersurface will be called the *yield surface.* The state under consideration is plastic or elastic according to whether the end point of the radius vector $\boldsymbol{\sigma}$ lies on the yield surface or in its interior. Since in the latter case a sufficiently small change of stress will not produce any change of plastic strain, we are interested only in the case where the end point of $\boldsymbol{\sigma}$ lies on the yield surface. In this case, an infinitesimal change of stress $d\boldsymbol{\sigma}$ constitutes unloading or loading or is neutral according to whether the vector $d\boldsymbol{\sigma}$ points from the extremity of $\boldsymbol{\sigma}$ towards the interior or the exterior of the yield surface or is tangential to it. It seems reasonable to assume that from any given plastic state a stress-free state can be reached by a suitable unloading process. This means that the origin must lie in the interior of the yield surface.

If awkward discontinuities in the stress-strain relations are to be avoided, we must be permitted to consider a neutral change of stress as a limiting case of either loading or unloading. In the following, this important condition will be called the *condition of continuity.*[1]

Let us choose a sequence of vectors $d\boldsymbol{\sigma}$ which are all directed towards the exterior of the yield surface but tend toward a limiting vector in the tangent plane of this surface. To each of these $d\boldsymbol{\sigma}$ there corresponds a uniquely determined change of plastic strain $d\boldsymbol{\epsilon}$. The condition of continuity requires that $d\boldsymbol{\epsilon} \to 0$ as $d\boldsymbol{\sigma}$ approaches the limiting vector.

A vector $d\boldsymbol{\sigma}$ directed towards the exterior of the yield surface may be considered as the sum of a vector $d\boldsymbol{\sigma}_n$ which is normal to the yield surface and a vector $d\boldsymbol{\sigma}_t$ which is tangent to this surface. Since the stress-strain law for loading is assumed to be linear in the components of the stress increment, the change in plastic strain produced by $d\boldsymbol{\sigma}$ may be found as the sum of the changes of plastic strain which $d\boldsymbol{\sigma}_n$ and $d\boldsymbol{\sigma}_t$ produce separately. On account of the condition of "continuity," however, $d\boldsymbol{\sigma}_t$ does not produce any change in plastic strain. Thus, the change of plastic strain $d\boldsymbol{\epsilon}$ associated with the stress increment $d\boldsymbol{\sigma}$ is seen to depend only on the normal component $d\boldsymbol{\sigma}_n$ of $d\boldsymbol{\sigma}$. This fact, together with the linearity of the relation between the increments of stress and plastic strain, shows that the direction of $d\boldsymbol{\epsilon}$ is completely independent of the direction of $d\boldsymbol{\sigma}$, while the magnitude of $d\boldsymbol{\epsilon}$ is proportional to that of $d\boldsymbol{\sigma}_n$.

III. THE CONDITION OF UNIQUENESS

The main result of the preceding section may be stated as follows. To each point $\boldsymbol{\sigma}$ of the yield surface a unit vector $\mathbf{n}$ may be associated such that the increment of plastic strain $d\boldsymbol{\epsilon}$ produced by any infinitesimal loading from $\boldsymbol{\sigma}$ has the direction of $\mathbf{n}$. We shall now investigate the relationship between the unit vectors $\mathbf{n}$ and the shape of the yield surface.

Let us consider a body made of the type of work-hardening material considered here and subject it to surface tractions which are varied in a specified manner, starting from zero. We now propose to study the following boundary value problem. Given

[1] Cf. G. H. Handelman, C. C. Lin, and W. Prager, Q. App. Math. 4, 397–407 (1947).

the instantaneous mechanical state throughout the body and a system of infinitesimal increments of the surface tractions, to find the corresponding stress increments throughout the body. It seems reasonable to demand that the solution of this boundary value problem be unique.[2] We shall investigate what restrictions this *uniqueness condition* imposes on the stress-strain law.

Let us assume that our boundary value problem admits two solutions. Using the geometrical representation of the preceding section, we represent the difference of the two stress increments at a given point of the body by $\Delta d\boldsymbol{\sigma}$, the difference of the increments of elastic strain by $\Delta d\mathbf{e}$, and the difference of the increments of plastic strain by $\Delta d\boldsymbol{\varepsilon}$. Now, the two solutions correspond to the same increments of surface tractions; moreover, the stress-strain law is linear in the increments of stress and strain. Using these facts and the divergence theorem, we may show that the integral of the scalar product

$$dI=\Delta d\boldsymbol{\sigma}\cdot(\Delta d\mathbf{e}+\Delta d\boldsymbol{\varepsilon}), \tag{1}$$

extended over the volume of the body under consideration, must vanish. Any stress-strain law which assures that the integrand dI is positive definite will therefore satisfy the condition of uniqueness.

Now, $\Delta d\mathbf{e}$ is related to $\Delta d\boldsymbol{\sigma}$ by Hooke's law and the scalar product $\Delta d\boldsymbol{\sigma}\cdot\Delta d\mathbf{e}$ is positive definite. As regards the scalar product $\Delta d\boldsymbol{\sigma}\cdot\Delta d\boldsymbol{\varepsilon}$, the following three cases must be discussed separately: (i) both solutions constitute loading at the point under consideration, (ii) one constitutes loading, the other unloading, and (iii) both solutions constitute unloading. In the third case $\Delta d\boldsymbol{\varepsilon}=0$, so that (i) is positive definite because $\Delta d\boldsymbol{\sigma}\cdot\Delta d\mathbf{e}$ is. In the other two cases, the order in which the two solutions are taken does not affect the sign of the product $\Delta d\boldsymbol{\sigma}\cdot\Delta d\boldsymbol{\varepsilon}$, because both $\Delta d\boldsymbol{\sigma}$ and $\Delta d\boldsymbol{\varepsilon}$ change sign when this order is reversed. We may therefore agree to order the two solutions so that the vector $\Delta d\boldsymbol{\sigma}$ points from the extremity of $\boldsymbol{\sigma}$ towards the exterior of the yield surface.

In the first case, $\Delta d\boldsymbol{\varepsilon}$ is related to $\Delta d\boldsymbol{\sigma}$ by the stress-strain law discussed in the preceding section. The direction of $\Delta d\boldsymbol{\varepsilon}$ is given by the unit vector $\mathbf{n}$ which depends only on the instantaneous state but not on the stress increment $\Delta d\boldsymbol{\sigma}$. It is easily seen that $\mathbf{n}$ must be the unit vector along the exterior normal of the yield surface, if the product $\Delta d\boldsymbol{\sigma}\cdot\Delta d\boldsymbol{\varepsilon}$ is to be non-negative for all vectors $\Delta d\boldsymbol{\sigma}$ which point towards the exterior of this surface. Moreover, in the second case, too, is the product $\Delta d\boldsymbol{\sigma}\cdot\Delta d\boldsymbol{\varepsilon}$ made non-negative by this choice of $\mathbf{n}$.

We are now in a position to explain why $d\sigma_x, \cdots, d\tau_{zx}\sqrt{2}$ and $d\epsilon_x, \cdots, d\gamma_{zx}/\sqrt{2}$ were taken as the components of $d\boldsymbol{\sigma}$ and $d\boldsymbol{\varepsilon}$. The direction of $\mathbf{n}$ has just been deduced from the postulate that the expression

$$d\sigma_x d\epsilon_x+d\sigma_y d\epsilon_y+d\sigma_z d\epsilon_z+d\tau_{xy}d\gamma_{xy}+d\tau_{yz}d\gamma_{yz}+d\tau_{zx}d\gamma_{zx}$$

be non-negative and the fact that this expression equals the scalar product $d\boldsymbol{\sigma}\cdot d\boldsymbol{\varepsilon}$. The same arguments would apply if $d\sigma_x, \cdots, \lambda d\tau_{zx}$, and $d\epsilon_x, \cdots, d\gamma_{zx}/\lambda$ were taken as the components of $d\boldsymbol{\sigma}$ and $d\boldsymbol{\varepsilon}$. The choice $\lambda=\sqrt{2}$ is motivated as follows. The statement that, for loading, $d\boldsymbol{\varepsilon}$ is directed along the exterior normal of the yield surface has physical significance only if it remains valid when the stress and the increment of plastic strain are referred to another set of rectangular axes x', y', z'. It can be shown that this condition is satisfied only if $\lambda=\sqrt{2}$.[3]

IV. THE CONDITIONS OF IRREVERSIBILITY AND CONSISTENCY

On account of the irreversible character of plastic deformation, work expended on plastic deformation cannot be reclaimed. This means that the work of the stresses on the change of plastic strain is positive whenever a change of plastic strain occurs. Formulated in the terms of our geometrical representation, this *condition of irreversibility*[4] requires that $\boldsymbol{\sigma}\cdot d\boldsymbol{\varepsilon}>0$ if $d\boldsymbol{\varepsilon}\neq 0$, i.e., that the radius vector to any point of the yield surface and the exterior normal of the yield surface at this point must make an acute angle. In other words, the yield surface, which always encloses the origin, must be convex with respect to the origin.

We have yet to specify the magnitude of the vector $d\boldsymbol{\varepsilon}$. This may be done as follows. Let

$$f(\sigma_x, \cdots, \tau_{zx}, \epsilon_x, \cdots, \gamma_{zx})=0 \tag{2}$$

define the dependence of the yield surface on the plastic strain. The desired stress-strain law furnishes the increment of plastic strain $d\boldsymbol{\varepsilon}$ which is produced by subjecting the stress of a given plastic state to a given increment $d\boldsymbol{\sigma}$. Since loading from a plastic state must again lead to a plastic state, the stress and plastic strain which exist after the infinitesimal changes $d\boldsymbol{\sigma}$ and $d\boldsymbol{\varepsilon}$ have taken place must still satisfy (2). Taking this *consistency condition* into account and expressing the geometrical results of the previous sections analytically, we obtain the following stress-strain law:

$$\begin{aligned} d\epsilon_x&=(dN/D)(\partial f/\partial\sigma_x), \ \cdots, \\ d\gamma_{zx}&=(dN/D)(\partial f/\partial\tau_{zx}), \end{aligned} \tag{3}$$

[2] Cf. E. Melan, Ing. Archiv 9, 116–126 (1938).

[3] Cf. H. Geiringer, *Fondements mathématiques de la théorie des corps plastiques.* (Mémorial Sci. Math.—Gauthier-Villars, Paris, 1937), fas. 86, pp. 19–22.

[4] Cf. P. Hodge and W. Prager, J. Math. Phys. 27, 1–10 (1948).

where

$$dN=(\partial f/\partial\sigma_x)d\sigma_x+\cdots+(\partial f/\partial\tau_{xz})d\tau_{xz}, \quad (4)$$

$$D=-(\partial f/\partial\sigma_x)(\partial f/\partial\epsilon_x) -\cdots-(\partial f/\partial\tau_{xz})(\partial f/\partial\gamma_{xz}). \quad (5)$$

The stress-strain law (3) is meant for loading from a plastic state ($dN>0, f=0$). Applied to a neutral change of stress ($dN=0, f=0$) it gives the correct result $d\epsilon_x=\cdots=d\gamma_{xz}=0$, thus fulfilling the condition of continuity. The condition of uniqueness is found to be satisfied when

$$D>0, \quad (6)$$

and the condition of irreversibility when

$$(\partial f/\partial\sigma_x)\sigma_x+\cdots(\partial f/\partial\tau_{xz})\tau_{xz}>0. \quad (7)$$

In spite of the generality of our basic assumptions, the form of the stress-strain law thus is subject to rather specific restrictions. A function which satisfies all these conditions, at least for sufficiently small plastic strains, is given by

$$f=(s_x-ce_x)^2+\cdots+2(\tau_{xz}-c\gamma_{xz}/2)^2-2k^2, \quad (8)$$

where

$$s_x=(2\sigma_x-\sigma_y-\sigma_z)/3,\ \cdots, \qquad e_x=(2\epsilon_x-\epsilon_y-\epsilon_z)/3,\ \cdots \quad (9)$$

are the normal components of the deviations of stress and plastic strain, k is the yield stress in pure shear for the first plastic deformation from the virgin state, and c is a constant which has the dimension of a stress. With this choice of f, the stress-strain law for plastic loading assumes the form

$$d\epsilon_x=(dN/2ck^2)(s_x-ce_x),\ \cdots, \qquad d\gamma_{xz}=(dN/ck^2)(\tau_{xz}-c\gamma_{xz}/2), \quad (10)$$

where

$$dN=(s_x-ce_x)ds_x+\cdots+2(\tau_{xz}-c\gamma_{xz}/2)d\tau_{xz}. \quad (11)$$

The yield condition (8) was first suggested by E. Reuss,[5] who was attracted to it because it can be made to represent the Bauschinger effect. In the geometric language used above, this function f may be said to represent an elliptic cylinder with generators normal to the plane $\sigma_x+\sigma_y+\sigma_z=0$. The position of the axis of this cylinder depends on the plastic strain, but the form and orientation of the cross section is independent of this strain.

V. INCREMENTAL *VERSUS* FINITE STRESS-STRAIN LAWS

The preceding example illustrates the general technique used in the discussion of stress-strain laws. Once the initial assumption is made as to what variables determine the mechanical state, the conditions of continuity, uniqueness, irreversibility, and consistency are used to extract as much information as is possible concerning the form of the stress-strain law. Since the particular initial assumption made above imposed rather far-reaching restrictions on the choice of the stress-strain law, one may wish to replace it by a less restrictive assumption. We may assume, for instance, that the mechanical state is determined by the stress and the strain history as defined by the *length of the strain path*, $\int|d\epsilon|$.[6] No matter what assumption is made regarding the variables which determine the mechanical state, the stress-strain laws obtained by this technique furnish the infinitesimal increment of strain when the mechanical state and the infinitesimal increment of the stress are known.

Given an initial plastic state and a finite increment of stress reached along a known stress path which avoids unloading, we can integrate the stress-strain law to obtain the corresponding finite increment of strain. As a rule, this will depend not only on the initial plastic state and the finite increment of stress but also on the entire stress path. In this case, we shall speak of an *incremental stress-strain law*. If, however, an initial plastic state and a finite increment of stress determine a unique increment of strain, the stress-strain law will be called *finite*.[7] In a stress analysis based on an incremental law, the effect of each infinitesimal load increment must be traced;[8] in a stress analysis based on a finite law, on the other hand, finite load increments can be considered provided they are known to avoid unloading. Therefore it is not surprising that finite stress-strain laws have been used in numerous investigations of plastic deformations of structural elements and machine parts.[9]

Unfortunately, it can be shown that no finite stress-strain law satisfies the condition of continuity.[1,10] As a rule, the validity of stress analyses based on finite stress-strain laws must therefore be questioned. This does not exclude the possibility, however, that in certain special cases such stress analyses may furnish correct or approximately correct results.

[5] E. Reuss, Abstracts, 4th Internat. Cong. App. Mech., Cambridge, 1934, p. 91; see also W. Prager, Zeits. f. angew. Math. Mech. 15, 76–80 (1935).

[6] Cf. D. M. Cunningham, E. G. Thomsen, and J. E. Dorn, Proc. Am. Soc. Test. Mat. 74, 546–553 (1947).

[7] Cf. K. H. Swainger, Phil. Mag. (7) 36, 443–473 (1945).

[8] Cf. R. Hill, E. H. Lee, and S. J. Tupper, Proc. Roy. Soc. London A191, 278–303 (1947); G. I. Taylor, Q. J. Mech. App. Math. 1, 103–124 (1948).

[9] Cf. N. M. Beliaev and A. K. Sinitsky, Izvestia Akad. Nauk, S.S.S.R. No. 2, 3–54 (1938); P. Locatelli, Ist. Lombardo Sci. Lett. Cl. Sci. Mat. Nat., Rend. (3) 4, 581–598 (1940); B. Finzi, Atti Ac. Sci. Torino, Cl. Sci. Fiz. Mat. Nat. 76, 222–238 (1941); F. P. Cozzone, J. Aero. Sci. 10, 137–151 (1943); A. A. Ilyushin, Prikladnaia Mat. Mekh. 7, 245–272 (1943); V. V. Sokolovsky, *ibid.* 7, 273–292 (1943); W. R. Osgood, J. Aero. Sci. 11, 213–226 (1944); E. Melan, Oesterreich. Ing. Arch. 1, 14–21 (1946).

[10] G. H. Handelman and W. Prager, Prikladnaia Mat. Mekh. 11, 291–292 (1947); W. Prager, J. App. Phys. 19, 540–543 (1948).

VI. STRUCTURAL STABILITY IN THE PLASTIC RANGE

As a rule, stress analyses based on incremental laws are rather cumbersome, because the deformation process must be treated as an infinite sequence of infinitesimal changes of state. There is an important group of problems, however, where it is sufficient to consider a single infinitesimal change of state. Indeed, the solution of a problem of structural stability in the plastic range requires only that a given configuration be compared to neighboring configurations which are obtained from the first by infinitesimal deformations. The difficulty usually experienced in the application of incremental stress-strain laws is therefore not encountered in stability problems. In spite of this fact, most investigations of the plastic buckling of plates and shells seem to have been based on finite laws.[11] Applied to the plastic buckling of a rectangular plate under edge thrusts, the two types of stress-strain laws yield clearly distinct values of the buckling stress, and the available experimental results agree best with the predictions furnished by finite laws.[12] This fact is sometimes interpreted as an indirect experimental verification of finite stress-strain laws. It is more likely, however, that the unsatisfactory agreement between the buckling stress found by experiment and that obtained from incremental stress-strain laws is due to the fact that the customary formulation of the buckling problem (adopted in all papers mentioned in reference 11) is incorrect.[13] The problem of column stability, for instance, is concerned with the determination of the smallest axial load for which the column can assume a bent, as well as a straight, equilibrium configuration. The usual treatment of column stability in the plastic range, however, is concerned with a more restricted problem, *viz.*, the determination of the smallest axial load under which the column can *pass* from the straight to a bent equilibrium configuration. In the elastic range, the difference between these two problems disappears, because the loads necessary to maintain an elastic system in a certain configuration depend only on this configuration and not on the manner in which it has been reached. In the plastic range, on the other hand, the strain history is of importance, and the two problems formulated above have distinct solutions. In the case of column stability, it turns out that the critical load is smaller for the first problem than for the second.

VII. STATICALLY DETERMINATE PROBLEMS IN THE SAINT VENANT-MISES THEORY

So far, this report on recent developments in the mathematical theory of plasticity has been concerned with work-hardening materials. The increased structural use of work-hardening aluminum alloys is doubtless responsible for the special attention which this branch of the theory has recently received. At the same time, however, there have been important developments in the theory of perfectly plastic materials. Within this field, the so-called statically determinate problems have probably received more attention than any other group of problems.

The torsion problem is the simplest of these; we shall review it briefly, because it lends itself well to the discussion of the typical features of all statically determinate problems.

The state of stress at a generic point P of the cross section of a prismatic or cylindrical bar in torsion is defined by the *shear stress vector*. The equation of equilibrium shows that this vector can be derived from a scalar *stress function*, the value of which depends on the position of the point P. The shear stress vector is obtained by rotating the gradient of the stress function in the cross-sectional plane through a right angle in the clockwise sense. At a generic point of the contour of the cross section, the shear stress vector must be tangent to the contour. Accordingly, the stress function must have a constant value along the contour, e.g., the value zero.

In the plastic zone of the cross section of an elastic-plastic bar in torsion, the shear stress vector must have a constant intensity equal to the yield stress in pure shear. The *stress surface*, whose ordinates represent the stress function, will therefore be a surface of constant slope. If the entire bar is plastic, the stress surface is completely determined as a surface of a given constant slope, with the contour of the cross section as directrix. The *fully plastic* stress distribution can therefore be determined without reference to the strain—hence the use of the term *statically determinate*.

The fully plastic stress distribution is of particular importance because it furnishes the maximum torque which the bar can sustain. In many other cases, too, the determination of the limit load of a structure made of a perfectly plastic material turns out to be a statically determinate problem. Thus, the determination of the safety factor of statically indeterminate trusses, beams, and frames with respect to the limit load may be simpler than the determination of the conventional safety factor

[11] *For the use of finite laws* cf., for instance, P. P. Bijlaard, Kon. Nederlandsche Akad. Wet., Proc. **41**, 468–480, 731–743; Mitteilungen Inst. Baustatik, E. T. H. Zurich, No. 21 (1947); A. A. Ilyushin, Prikladnaia Mat. Mekh. **8**, 337–360 (1944), *ibid.* **10**, 623–638 (1946). *For the use of incremental laws* cf. I. P. Kuntze, Comptes Rendus Ac. Sci. U.R.S.S. **55**, 387–389 (1947); G. H. Handelman and W. Prager, N.A.C.A. Tech. Note. No. 1530 (1948).

[12] C. F. Kollbrunner, Mitteilungen Inst. Baustatik, E.T.H. Zurich, No. 17 (1946); C. F. Kollbrunner and G. Herrmann, *ibid.* No. 20 (1947).

[13] Cf. F. R. Shanley, J. Ae. Sci. **14**, 261–267 (1947); W. Prager, *ibid.* **15**, 337–341 (1948).

with respect to the limit of proportionality. This new method of *limit design* was first advocated by Kist[14] and has received much attention in recent years.[15]

Returning to the torsion problem, let us briefly consider the behavior of the bar under a torque which is smaller than the limit torque but sufficiently large to produce yielding in some portion of the cross section. The stress function can then be determined experimentally by the well-known soap film-sand hill analogy.[16] The analytical determination of the stress function for a given cross section proves exceedingly difficult,[17] but relaxation methods have been used successfully to obtain numerical results.[18]

In view of the mathematical difficulties encountered in treating the problem of elastic-plastic torsion, it is not surprising that, with one notable exception,[19] the problem of elastic-plastic plane strain has been treated only in trivial cases of cylindrical symmetry. The following remark[20] proves useful in the qualitative discussion of such problems. In elastic plane strain the stress function of Airy is biharmonic and can therefore be represented by the small deflections of a thin plate which is bent by forces and couples applied along its contour. This analogy corresponds to the soap film analogy for elastic torsion. For elastic-plastic plane strain the deflections of this plate must be restricted by a suitable rigid surface in the same manner in which, in the torsion problem, the deflections of the soap film are restricted by a moulding of the sand hill. The shape of this restricting surface is obtained from the *fully plastic* stress distribution.

In plane strain, as in torsion, the fully plastic stress distribution is statically determinate if the boundary conditions involve only stresses and not displacements. By means of the yield condition, the three stress components σ_x, σ_y, and τ_{xy} can be expressed in terms of two parameters, the mean pressure, and the angle which the principal axes of stress form with the coordinate axes. The equilibrium conditions then furnish a hyperbolic system of first-order partial differential equations for these two parameters. The analytical integration of this system has been discussed by Christianovich[21] and Geiringer.[22] Schemes of numerical integration based on the method of characteristics have been used successfully, in particular by Russian authors.[23] Ever since Hencky's original investigation,[24] the geometry of the slip lines (characteristics) has received much attention. For the drawing of the net of slip lines the following remark proves useful:[25] the distance between two neighboring slip lines of the same family is a linear function of the arc length of these lines. Among the less well-known results concerning the geometry of the slip lines, we mention the following:[26] the evolutes of one family of slip lines, together with their normal trajectories, form again a net of slip lines.

For the torsion problem, it is well known that the fully plastic stress distribution cannot be continuous. In other terms, the surface of the sand hill does not have a continuously turning tangent plane but has edges or ridges. At a generic point of such a line of discontinuity in the cross-sectional plane, the normal component of the shear stress vector has the same value on both sides of the line of discontinuity but the tangential component varies in a discontinuous manner.

Similar discontinuities of stress arise in problems of plane plastic strain.[27] At a point where the line of discontinuity is tangent to the y axis, for instance, σ_x and τ_{xy} must be continuous across the line of discontinuity if equilibrium is to be preserved, but σ_y may vary in a discontinuous manner.

In the torsion problem, the fully plastic stress distribution is known to exist and to be unique. For statically determinate problems of plane plastic strain, the uniqueness of the fully plastic stress distribution can be deduced from recent more general results of Markov.[28] Though Markov's uniqueness proof is based on the assumption of a continuous stress distribution, it can easily be modified so as to include discontinuities of stress. As a matter of fact, fully plastic stress distributions without discontinuities are seldom encountered in practical problems.

VIII. AN IMPORTANT PROBLEM OF THE THEORY OF PLASTICITY

Any stress analysis aims, of course, at the prediction of the stresses produced by given loads in the

[14] In his inaugural lecture at the Technical University of Delft, Oct. 2, 1917.
[15] Cf. the bibliography in Van Den Broek, *Theory of Limit Design* (John Wiley and Sons, Inc., New York, 1948).
[16] A. Nadai, Zeits. f. angew. Math. Mech. **3**, 442–454 (1923).
[17] Cf. E. Trefftz, Zeits. f. angew. Math. Mech. **5**, 64–73 (1925); L. A. Galin, Prikladnaia Mat. Mekh. **8**, 307–322 (1944).
[18] Cf. D. G. Christopherson, Trans. A.S.M.W. **62**, A1–A4 (1940). F. S. Shaw, Australian Council for Aeronautics, Rep. ACA–11 (1944).
[19] L. A. Galin, Prikladnaia Mat. Mekh. **10**, 365–386 (1946).
[20] W. Prager, *Theory of Plasticity* (mimeographed lecture notes, Brown University, Providence, Rhode Island, 1942), p. 148.
[21] C. A. Christianovich, Mat. Sbornik **1**, 511–534 (1936).
[22] See reference 3, Chap. II.
[23] Cf. V. V. Sokolovsky, J. App. Mech. **13**, A1–A10 (1946); see also the same author's *Theory of Plasticity*, Izd. Ak. Nauk S.S.S.R., Moscow and Leningrad, 1946.
[24] H. Hencky, Zeits. f. angew. Math. Mech. **3**, 241–251 (1923).
[25] W. Prager, Rev. Fac. Sci. Istanbul **4**, 22–24 (1939).
[26] I. Kapuano, Rev. Fac. Sci. Istanbul (A) **6**, 36–39 (1941); see also *ibid.* (A) **9**, 35–60 (1944).
[27] W. Prager, *R. Courant Anniversary Volume* (Interscience Publishers, Inc., New York, 1948), pp. 289–300.
[28] A. A. Markov, Prikladnaia Mat. Mekh. **11**, 339–350 (1947).

structure under consideration. Now, almost any structure has to support a variety of loads; as a rule, upper and lower limits may be assigned for the intensity of each of these loads, but the manner in which these loads occur jointly or follow each other is not specified. For an elastic structure this randomness of the loads does not cause any difficulty; the effects of each load can be studied separately and the most dangerous combination of loads can then be determined for any given point of the structure. For an elastic-plastic structure, however, the principle of superposition is not valid, and the stresses produced by a given combination of loads depend on the precise program of loading. On the other hand, structures with a precise loading program are not often encountered in engineering. Under these circumstances, the following problem is of particular importance. Given a structure under the influence of loads each of which is varying in a random manner between given extreme values. Will the structure eventually "shake down" to a state of residual stress such that all further variations of the loads between the given extremes are supported in a purely elastic manner and, if so, is this state of residual stress unique? For trusses made of perfectly plastic or work-hardening materials this problem has been discussed by Bleich[29] and Melan,[30] who established rather general conditions for the occurrence of shake-down. From their examples it follows that in many cases the shake-down state of residual stress is independent of the sequence of loads, but depends only on the extremes between which the individual loads are allowed to vary. In an attempt at generalizing these results one might, for instance, study a cylindrical bar under combined tension and torsion such that the axial force and the torque vary independently, each between given limits. If it should turn out that in such more general cases, too, the shake-down state of residual stress is independent of the precise loading program, this result would greatly increase the practical usefulness of the mathematical theory of plasticity.

[29] F. Bleich, Ossature Metall. **3**, 93-105 (1934).
[30] E. Melan, S. B. Akad. Wiss. Wien **1936**, 195–218.

Contributed Original Research

Some Properties of Tubular Electron Beams

NELSON WAX*
Bell Telephone Laboratories, Inc., New York, New York

(Received May 6, 1948)

Approximate expressions are obtained for the potential distribution, maximum current density, and beam spread of tubular electron beams of finite thickness. Comparisons are made with previously published results for very thin tubular beams and solid cylindrical beams.

I. INTRODUCTION

THE increasing interest in electronic devices in which an electron beam is allowed to interact with radiofrequency fields has caused a number of workers to consider the use of tubular electron beams, instead of the usual "solid" cylindrical beams. There may be a number of advantages in using tubular beams; the maximum current density is greater, the beam spread is less (at least for current densities much less than the maximum), and the electromagnetic fields may be more uniform over the cross section of the tubular beam than they are for the cylindrical beam.

Tubular electron beams of negligible thickness have been treated by Haeff;[1] the method used, which is due to Nergaard, is inapplicable to beams of finite thickness. Some of the properties of tubular beams of finite thickness are considered in this paper: the radial potential distribution, the maximum current density, and the beam spread. The problems of beam formation and electron motion during the initial period of acceleration are ignored here.

Some attention is given to the case where the inner conductor is absent, not only for its practical importance, but also because it serves as a relatively simple example of the somewhat more general results obtained.

The following assumptions are made throughout this work, except where noted:

(1) The system of electrodes, as well as the tubular electron beam, is axially symmetric.

(2) All electrons are emitted from a source with zero initial velocity; the source is at zero potential.

(3) The electrons, after the beam is formed, are constrained to travel in straight paths parallel to the axis, with uniform velocities.** This implies that the axial and azimuthal variations in potential can be neglected. (This assumption is modified slightly in the section on beam spread.)

(4) The electron velocities are very much less than the velocity of light; the applied differences of potential are never more than a few kilovolts therefore, and the magnetic effects of the moving electrons can be neglected.

(5) The current density is constant over the cross section of the beam for a fixed geometry and set of potentials.

Figure 1 gives a schematic representation of the system. M.k.s. units are used.

II. RADIAL POTENTIAL DISTRIBUTION AND CURRENT DENSITY

V_1, V_2, V_4 and r_1, r_2, r_3, r_4 are taken as known;*** the problem is to find the potential distribution and current density within the beam. The method used is to assume that the current density is known and obtain an approximate solution of the differential equation for the radial potential distribution using the two initial conditions at the inner boundary of the beam. A further condition at the outer boundary of the beam is then sufficient to determine the current density.

The potential distribution within the beam obeys the Poisson equation

$$\nabla^2 V = -\rho/\epsilon_0, \tag{1}$$

* Now at the University of Illinois, Urbana, Illinois.

[1] A. V. Haeff, Proc. I.R.E. 27, 586 (1939).

** One way of satisfying this assumption physically, to a good degree of approximation, is to use a system which is completely immersed in a very strong magnetic field. Another occurs in the absence of a magnetic field; the electron flow is approximately parallel at the plane where the beam cross section is a minimum. A similar assumption has been made by other workers. (See references 1 and 2).

*** In a practical case V_2 would not be known in advance, V_2 is determined, however, (in terms of the other potentials and the geometry) when the beam current is a maximum (Eq. (21)). In applying the results of this paper one should first calculate the maximum current density (using Eqs. (21), (14), and (15)); the potential distribution within the beam for current densities which are given fractions of the maximum can then be obtained by means of Eqs. (14), (15), and (9).

where V is the potential, ρ is the space charge density at any point within the beam, and ϵ_0 is the dielectric constant of free space $=8.854\times10^{-12}$ farad per meter.

The assumptions made in the introduction allow one to write Eq. (1) in cylindrical coordinates as

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right)=\frac{i}{\epsilon_0(2\eta V)^{\frac{1}{2}}}, \tag{2}$$

where the relation $-\rho v=i$ has been used. v is the velocity $=(2\eta V)^{\frac{1}{2}}$ and i is the current density. η is the ratio of the charge of the electron to its mass $=1.769\times10^{11}$ coulombs/kg.

It should be noted that $V\geqq0$ holds within the beam, since there can be no electron flow in regions where the potential is less than that of the source.

The initial conditions are that the potential and the electric field be continuous at the inner boundary of the beam.

$$V=V_2,\qquad \frac{dV}{dr}=\frac{V_2-V_1}{\log\frac{r_2}{r_1}}\cdot\frac{1}{r_2} \tag{3}$$

at $r=r_2$. Natural logarithms are used throughout.

It is convenient to introduce the dimensionless quantities

$$x=\log\frac{r}{r_2}, \tag{4}$$

$$\Phi=\frac{V-V_2}{V_4-V_2}, \tag{5}$$

$$p=\frac{ir_2^2}{\epsilon_0(2\eta)^{\frac{1}{2}}\alpha V_2^{\frac{3}{2}}}, \tag{6}$$

with $\alpha=V_4/V_2-1$.

Equation (2) becomes

$$\frac{d^2\Phi}{dx^2}=\frac{pe^{2x}}{(1+\alpha\Phi)^{\frac{1}{2}}} \tag{7}$$

with $e=2.718\cdots$.

The initial conditions given by Eq. (3) are, at $x=0$,

$$\Phi=0,\qquad d\Phi/dx=a_0=\beta/\alpha x_1 \tag{8}$$

with

$$\beta=(V_1/V_2)-1 \quad\text{and}\quad x_1=\log(r_1/r_2).$$

A power series solution of Eq. (7) may be obtained in the parameter p, subject to the conditions given above. This is

$$\Phi=a_0x+\sum_{n=1}^{\infty}a_np^n \tag{9}$$

with the coefficients given by

$$a_n=(-\alpha)^{n-1}[a_{n0}(x)-\gamma a_{n1}(x)], \tag{10}$$

where

$$\gamma=\alpha a_0=(\beta/x_1).$$

The values of the functions $a_{10}(x)$, $a_{11}(x)$, etc. and their first derivatives are listed in Table I. The mathematical details of the expansion, and a comparison with numerical solutions of Eq. (7) are given in the Appendix. The first four terms (terms including p^3) of the series provide a fairly good approximation to the solution of Eq. (7) even when the current density has its maximum value. It was assumed, in obtaining the coefficients a_{n1} that $|\gamma|\ll1$, and $x\leqq1$. The treatment given here is thus restricted to cases where the electric field is small at the inner beam boundary and to ratios of outer to inner beam radii no greater than 2.718.

$a_0=0$ when there is no inner conductor. Equation (9) then takes the somewhat simpler form

$$\Phi=\sum_{n=1}^{\infty}(-\alpha)^{n-1}a_{n0}(x)p^n. \tag{11}$$

The condition that the electric field be continuous at the outer beam boundary allows one to determine p.

At $r=r_3$

$$\frac{dV}{dr}=\frac{V_4-V_3}{\log(r_4/r_3)}\cdot\frac{1}{r_3} \tag{12}$$

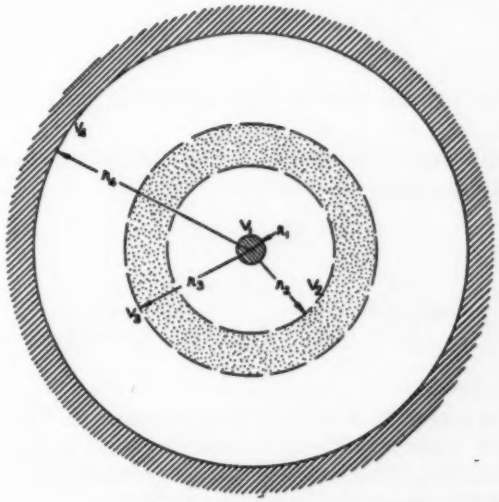


FIG. 1. A representation of the geometry and potentials is shown here. The electron motion is perpendicular to the plane of the paper.

TABLE I.

x	$a_{10}(x)$	$a_{10}'(x)$	$a_{11}(x)$	$a_{11}'(x)$	$a_{20}(x)$	$a_{20}'(x)$
0	0	0	0	0	0	0
0.1	5.351×10^{-3}	1.107×10^{-1}	9.225×10^{-5}	2.860×10^{-3}	1.725×10^{-6}	1.025×10^{-4}
0.2	2.296×10^{-2}	2.459×10^{-1}	8.183×10^{-4}	1.311×10^{-2}	4.612×10^{-5}	1.003×10^{-3}
0.3	5.553×10^{-2}	4.111×10^{-1}	3.065×10^{-3}	3.389×10^{-2}	2.778×10^{-4}	4.174×10^{-3}
0.4	1.064×10^{-1}	6.128×10^{-1}	8.085×10^{-3}	6.936×10^{-2}	1.042×10^{-3}	1.226×10^{-2}
0.5	1.796×10^{-1}	8.591×10^{-1}	1.761×10^{-2}	1.250×10^{-1}	3.040×10^{-3}	2.977×10^{-2}
0.6	2.800×10^{-1}	1.160	3.399×10^{-2}	2.080×10^{-1}	7.556×10^{-3}	6.422×10^{-2}
0.7	4.138×10^{-1}	1.528	6.043×10^{-2}	3.278×10^{-1}	1.685×10^{-2}	1.278×10^{-1}
0.8	5.883×10^{-1}	1.977	1.012×10^{-1}	4.965×10^{-1}	3.473×10^{-2}	2.401×10^{-1}
0.9	8.124×10^{-1}	2.525	1.619×10^{-1}	7.300×10^{-1}	6.750×10^{-2}	4.319×10^{-1}
1.0	1.097	3.195	2.500×10^{-1}	1.049	1.253×10^{-1}	7.513×10^{-1}

x	$a_{21}(x)$	$a_{21}'(x)$	$a_{30}(x)$	$a_{30}'(x)$	$a_{31}(x)$	$a_{31}'(x)$
0	0	0	0	0	0	0
0.1	2.499×10^{-7}	1.295×10^{-5}	4.424×10^{-9}	2.764×10^{-7}	9.068×10^{-10}	6.521×10^{-8}
0.2	9.622×10^{-6}	2.613×10^{-4}	3.630×10^{-7}	1.180×10^{-5}	1.496×10^{-7}	5.631×10^{-6}
0.3	8.823×10^{-5}	1.639×10^{-3}	5.327×10^{-6}	1.202×10^{-4}	3.340×10^{-5}	8.698×10^{-5}
0.4	4.499×10^{-4}	6.506×10^{-3}	3.878×10^{-5}	6.833×10^{-4}	3.289×10^{-5}	6.663×10^{-4}
0.5	1.677×10^{-3}	2.003×10^{-2}	1.928×10^{-4}	2.829×10^{-3}	2.072×10^{-4}	3.485×10^{-3}
0.6	5.081×10^{-3}	5.254×10^{-2}	7.544×10^{-4}	9.605×10^{-3}	9.867×10^{-4}	1.434×10^{-2}
0.7	1.344×10^{-2}	1.236×10^{-1}	2.507×10^{-3}	2.849×10^{-2}	3.879×10^{-3}	5.014×10^{-2}
0.8	3.224×10^{-2}	2.686×10^{-1}	7.408×10^{-3}	7.669×10^{-2}	1.327×10^{-2}	1.557×10^{-1}
0.9	7.167×10^{-2}	5.502×10^{-1}	2.003×10^{-2}	1.919×10^{-1}	4.091×10^{-2}	3.916×10^{-1}
1.0	1.503×10^{-1}	1.076	5.059×10^{-2}	4.539×10^{-1}	1.163×10^{-1}	1.174

or, at $x=x_3$,

$$\Phi+(x_4-x_3)(d\Phi/dx)=1. \tag{13}$$

When Eqs. (13) and Eq. (9) are combined, p may be obtained by the method of successive approximations. The first approximation is

$$p_1=(1/A_{10})[1-a_0x_4+\gamma(A_{11}/A_{10})] \tag{14}$$

and the second approximation is

$$p_2=p_1\left[1-p_1\frac{A_2+p_1A_3}{A_1+2p_1A_2+3p_1^2A_3}\right]. \tag{15}$$

The following notation has been used:

$$\begin{aligned} A_{lm}&=a_{lm}(x_3)+(x_4-x_3)a_{lm}'(x_3),\\ A_l&=(-\alpha)^{l-1}[A_{l0}-\gamma A_{l1}]. \end{aligned} \tag{16}$$

The primes denote differentiation with respect to x.

Equation (14) and Eq. (15) become, when the inner conductor is absent,

$$p_1=1/A_{10} \tag{17}$$

and

$$p_2=p_1\left[1+p_1\frac{\alpha A_{20}-\alpha^2p_1A_{30}}{A_{10}-2p_1\alpha A_{20}+3p_1^2\alpha^2A_{30}}\right], \tag{18}$$

respectively.

p is greatest, when $a_0=0$, if $x_4-x_3=0$. This is in accord with results obtained for solid cylindrical beams, the total current being greatest when the beam fills the entire enclosing tube.[2]

[2] L. P. Smith and P. L. Hartman, J. App. Phys. 11, 220–229 (1940).

Figure 2 shows Φ as a function of x, for various values of p, $a_0=0$ and $\alpha=2$. Figure 3 shows Φ as a function of x for three values of a_0, and fixed p and α. These curves were obtained from numerical solutions of Eq. (7).

III. MAXIMUM CURRENT DENSITY

The quantity J is defined by

$$J=\alpha pV_2^{\frac{1}{2}}; \tag{19}$$

it is proportional to the current density (see Eq. (6)). The condition that J be a maximum as a function of V_2 is $(dJ/dV_2)=0$, and this leads to

$$V_2(dp/dV_2)=p[(2-\alpha)/2\alpha]. \tag{20}$$

Equation (20) and Eq. (14) yield

$$\beta_{\max J}=\frac{2(x_4-x_1)+[x_1-2(A_{11}/A_{10})]\alpha_{\max J}}{x_4+(A_{11}/A_{10})(2+\alpha_{\max J})} \tag{21}$$

as a first approximation to the condition for maximum current density.

It is instructive to consider first the case where the inner conductor is absent.

p_1 is independent of V_2, when $a_0=0$, and $\alpha_{\max J}=2$ therefore, or

$$\left(\frac{V_4}{V_2}\right)_{\max J}\doteq 3. \tag{22}$$

This result may also be obtained from Eq. (21) by letting

$$r_1/r_2\to 0 (x_1\to-\infty).$$

With no inner conductor the potential at the inner beam boundary is $\frac{1}{3}$ that of the potential of the outer conductor for maximum current density. This indicates that an appreciable range of electron velocities exists within the beam when one is operating at, or near, maximum current density.

Haeff[1] has shown, for thin beams, "that when the current in the beam is increased to its maximum value, the space potential at the beam will decrease to one-third of its original value (i.e., for the condition when no current flowed)." Equation (22) is in agreement with this statement, provided one takes V_2 as being his "space potential at the beam." His result is subject to some modification, however, when one considers thick beams with an inner conductor present. This may be seen by means of an example. Let $V_4=V_1(\beta=\alpha)$. According to Haeff $\alpha_{maxJ}=2$, then. The equation which determines α_{maxJ} is, from Eq. (21),

$$\frac{A_{11}}{A_{10}}\alpha^2{}_{maxJ}+\left(x_4-x_1+4\frac{A_{11}}{A_{10}}\right)\alpha_{maxJ}=2(x_4-x_1) \quad (23)$$

and it is only when $|A_{11}/A_{10}|\ll(x_4-x_1)$ that $\alpha_{maxJ}=2$ is a good approximation.

Figure 4 shows β_{maxJ} as a function of α_{maxJ} for three values of r_1/r_2. The potential distributions shown in Fig. 2 are for the conditions of maximum current density. The maximum current density is

$$i_{max}=\frac{\epsilon_0(2\eta)^{\frac{1}{2}}J_{max}}{r_2{}^2}. \quad (24)$$

The maximum total current is

$$I_{max}=\pi(r_3{}^2-r_2{}^2)i_{max}$$
$$=3.31\times10^{-5}a_{10}'(x_3)\frac{\alpha_{maxJ}}{(\alpha_{maxJ}+1)^{\frac{1}{2}}}p_{maxJ}V_4{}^{\frac{3}{2}}. \quad (25)$$

As a numerical example, let $\alpha=\beta$, $x_3=0.2$, $r_1/r_2=\frac{1}{10}$, $r_4/r_3=10$. Then $\alpha_{maxJ}=1.88$ (from Eq. (23)). $p_{1maxJ}=3.47$, $p_{2maxJ}=3.52$ (the second approximation being within 2 percent of the first, here) and Eq. (25) becomes

$$I_{max}=1.10\times10^{-5}V_4{}^{\frac{3}{2}} \quad (26)$$

where I_{max} is in amperes and V_4 is in volts.

IV. BEAM SPREAD

Beam spread has engaged the attention of a number of workers. The method of analysis used here follows that of Smith and Hartman,[2] who pointed out that previous calculations were in error because the variations in velocity within the beam had been neglected.

One assumes in this section that most of the electrons continue to travel parallel to the axis, held together by the constraints imposed upon them. A few electrons, at the outer edge of the beam, are considered free to travel, however, to the outer cylinder under the action of the electric field alone, which exists between the beam and the outer cylinder.

The equation of motion of an electron, which was initially at the outer edge of the beam, is

$$\frac{d^2r}{dt^2}=\eta\frac{V_4-V_3}{\log(r_4/r_3)}\cdot\frac{1}{r}. \quad (27)$$

It can be integrated

$$\left(\frac{dr}{dt}\right)^2=2\eta\frac{V_4-V_3}{\log(r_4/r_3)}\log\frac{r}{r_3}, \quad (28)$$

assuming that its initial radial velocity was zero. The axial velocity is unchanged throughout the course of the motion and is given by

$$dz/dt=(2\eta V_3)^{\frac{1}{2}}. \quad (29)$$

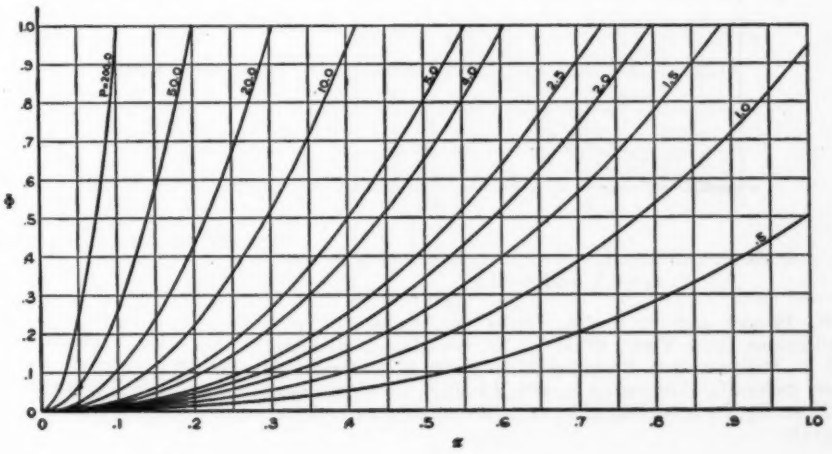


FIG. 2. Φ shown as a function of x, for various values of p. $a_0=0$ and $\alpha=2$.

Using Eq. (29) in Eq. (28), rearranging, and integrating, one obtains the electron trajectory

$$z=\frac{[1+\alpha\Phi(x_3)]^{\frac{1}{2}}(x_4-x_3)^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}[1-\Phi(x_3)]^{\frac{1}{2}}}\int_{r_2}^{r}\frac{dr}{\log^{\frac{1}{2}}(r/r_2)}, \quad (30)$$

or

$$\frac{z}{r_2}=\frac{2[1+\alpha\Phi(x_3)]^{\frac{1}{2}}(x_4-x_3)^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}[1-\Phi(x_3)]^{\frac{1}{2}}}\int_0^{\log^{\frac{1}{2}}(r/r_2)}\exp(\xi^2)d\xi. \quad (31)$$

The latter integral has been tabulated,[3] over a useful range.

The ratio of the distances required for a hollow beam, and a solid cylindrical beam, to diverge the same amount, is shown in Fig. 5, as a function of the fraction of the maximum total current. The data for the solid beam have been taken from the paper by Smith and Hartman.[2] It is noteworthy that the beam spread is very nearly the same in both cases. Thus for the same current density the hollow beam will diverge less than the solid beam.

The author thanks Miss M. E. Moore, of the Bell Telephone Laboratories, Inc., for her efficient performance of all the calculations, and her care in checking the algebraic details.

APPENDIX

Equation (7) is

$$\frac{d^2\Phi}{dx^2}=\frac{pe^{2x}}{(1+\alpha\Phi)^{\frac{1}{2}}}. \quad (A1)$$

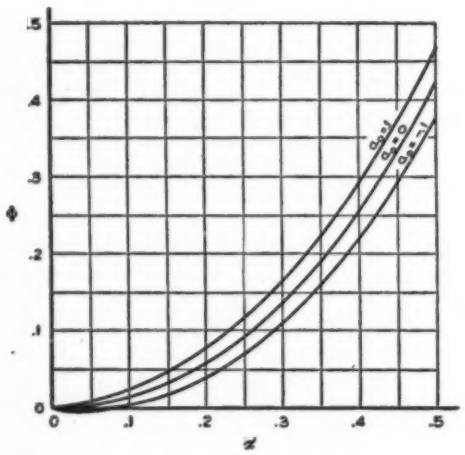


FIG. 3. Φ shown as a function of x for three values of a_0. $\alpha=2$, $x_3=0.5$, and $p=2.5$.

[3] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943), p. 32. See also A. Fletcher, J. C. P. Miller, and L. Rosenhead, *An Index of Mathematical Tables* (Scientific Computing Service, Limited, London, 1946), p. 219 *et. seq.* for a bibliography of tables of this, and related, integrals.

One assumes an expansion of the form

$$\Phi=a_0x+\sum_{n=1}^{\infty}a_np^n \quad (A2)$$

substitutes (A2) in (A1), expands by the binomial theorem, and equates coefficients of the same power of p. One is thus led to the following set of differential equations:

$$a_1''=\frac{e^{2x}}{(1+\gamma x)^{\frac{1}{2}}}, \quad (A3)$$

$$a_2''=-\frac{\alpha}{2}a_1\frac{e^{2x}}{(1+\gamma x)^{\frac{1}{2}}}, \quad (A4)$$

$$a_3''=\frac{\alpha}{2}(\tfrac{3}{4}\alpha a_1^2-a_2)\frac{e^{2x}}{(1+\gamma x)^{\frac{1}{2}}}. \quad (A5)$$

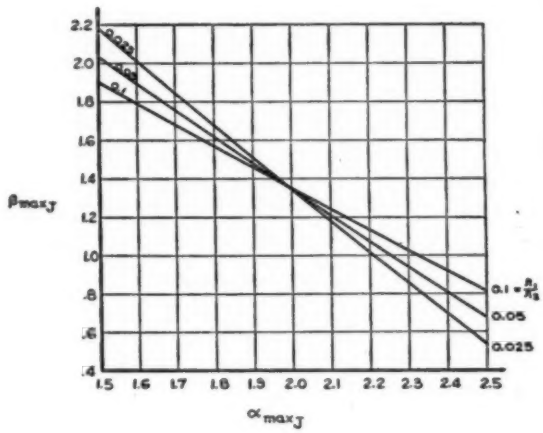


FIG. 4. $\beta_{\max J}$ shown as a function of $\alpha_{\max J}$ for three values of r_1/r_2. $x_4=2$, $x_3=0.5$.

The primes denote differentiation with respect to x.

The boundary conditions $\Phi=0$, $\Phi'=a_0$, at $x=0$ give $a_n=a_n'=0$ at $x=0$ $(n\neq 0)$.

If one neglects quadratic and higher order terms in γ, then

$$a_1''=e^{2x}-\gamma\frac{xe^{2x}}{2}=a_{10}''-\gamma a_{11}'', \quad (A6)$$

$$a_2''=-\frac{\alpha e^{2x}}{2}[a_{10}-\gamma(a_{11}+\tfrac{3}{2}xa_{10})]$$
$$=-\alpha(a_{20}''-\gamma a_{21}''), \quad (A7)$$

$$a_3''=\left[\frac{3}{8}\frac{\alpha^2a_1^2n}{(1+\gamma x)^{5/2}}-\frac{\alpha}{2}\frac{\alpha_2}{(1+\gamma x)^{\frac{1}{2}}}\right]e^{2x}$$
$$=\alpha^2(a_{30}''-\gamma a_{31}''). \quad (A8)$$

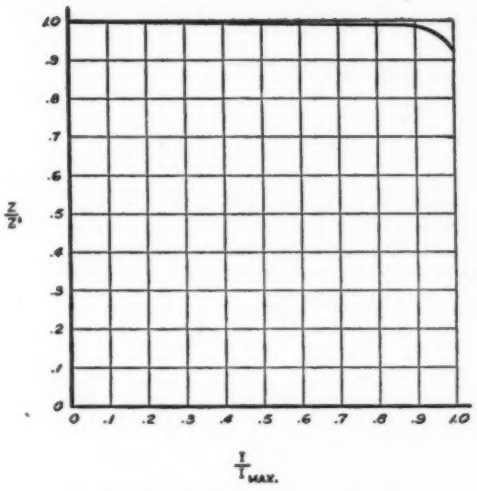


FIG. 5. The ratio of the distance required for a hollow beam, z, to the distance required for a solid cylindrical beam, z', to diverge the same amount is shown as a function of the fraction of the total maximum current. $r_4/r_3=10$ in both cases. $x_3=0.2$, for the hollow beam. The maximum total currents are within a few percent of each other and were taken as equal in order that $z/z'=1$ at $I/I_{max}=0$.

The functions a_{lm}'' may be integrated analytically with $a_{lm}=a_{lm}'=0$ at $x=0$. The first few are

$$a_{10}'=\frac{e^{2x}-1}{2}, \tag{A9}$$

$$a_{10}=\frac{e^{2x}-2x-1}{4}, \tag{A10}$$

$$a_{11}'=\frac{xe^{2x}}{4}-\frac{a_{10}'}{4}, \tag{A11}$$

$$a_{11}=\frac{a_{11}'}{2}-\frac{a_{10}}{4}, \tag{A12}$$

$$a_{20}'=\frac{1}{8}\left[\frac{e^{4x}-1}{4}-4a_{11}'-a_{10}'\right], \tag{A13}$$

$$a_{20}=\frac{1}{8}\left[\frac{e^{4x}-4x-1}{16}-4a_{11}-a_{10}\right]. \tag{A14}$$

The functions a_{21}, a_{30}, and a_{31} have also been obtained, but are not given here, as they are complicated. Their values are given in Table I, as well as those for the functions listed above.

Comparisons of the values obtained from Eq. (A2) and those obtained by numerical integration of Eq. (A1) have been made. The error, in neglecting terms higher than those in p^3, is always less than 7 percent, for $p=p_{maxJ}$, $a_0=0$, and $x_4=x_3$. p is greatest, when $a_0=0$, for $x_4=x_3$, as was noted under Eq. (18). The convergence improves quite rapidly as either of the inequalities $p<p_{maxJ}$ or $x_4>x_3$ becomes greater. Thus, for the curves shown in Fig. 3, the error was less than 3 percent, when terms higher than those in p^2, were neglected.

Only small values of γ have been considered because of the analytical difficulties in treating the case where γ is unrestricted in value; physically, the latter would certainly be of interest. One can indicate the nature of the difficulties involved when there are no restrictions on γ by considering Eq. (A5). Integrating it once yields an integral of the same form as that shown in Eq. (31); a_1 is given by a second integration, a_2 by two further integrations involving a_1 (Eq. (A6)), and the resulting expressions soon become unmanageable.

Equation (A5) has been integrated numerically from $x=0$ to $x=1$, for a number of values of γ. It was found that for $\gamma\leqq 0.5$ the error in using the linear approximation for $a_1'(a_1'=a_{10}'-\gamma a_{11}')$ is less than 5 percent, the error decreasing as γ decreases.

A New Solution to the Problem of Vertical Angle-of-Arrival of Radio Waves*

E. W. HAMLIN, P. A. SEAY,** W. E. GORDON***
University of Texas, Austin, Texas

(Received May 17, 1948)

A practical mathematical solution is presented, which separates the direct and reflected components of microwave radiation when a single pair of plane waves is received. Very promising operational applications are noted when it is desired to determine elevation angles precisely, continuously, and instantaneously. The method is well suited to the measurement of very small elevation angles. The results of sample tests of the solution warrant further study in this direction.

I. INTRODUCTION

THIS laboratory has been engaged in the study of the angle-of-arrival of microwave radiation for a period of about two years. The phase difference measurement equipment[1] was developed to determine the angle-of-arrival and potentially is capable of a very high degree of accuracy. This potential is realized when only one plane wave is received. However, this situation is rarely encountered at angles close to grazing.[2-4]

In field studies for small elevations a reflected component is invariably found. Though this component may be very weak or variable compared to the direct wave, its presence complicates the phase front,[5] and makes the desired angle difficult to determine.

To continue the study of the angle-of-arrival, the laboratory was confronted with two choices: (1) an attempt at eliminating the reflected component might be made by carefully selecting a path and properly siting the receiver, or (2) accept the reflected component at the receiver, and eliminate its effect in the equipment or in the calculations which might be applied to the data.

The second choice is far more desirable from the point of view of continuing the basic studies, since it does not impose rigid restrictions as to locations and sitings. This is a more natural choice and, further, of importance to the services. If successfully accomplished it would provide a highly accurate means of measuring elevation angles. A practical mathematical solution which separates the direct and reflected components is presented in this report. It should be especially noted that the method does not break down at very low elevations, but on the contrary was primarily developed for use with elevation angles less than one degree.

The solution, while somewhat cumbersome for manual calculations, could be handled with little difficulty by an electronic computer to provide precise, continuous, and instantaneous angles of elevation.

The data required to solve for the angle-of-arrival is obtained from two phase differences and two signal strength ratios from a set of three vertically spaced antennas. With the present two-antenna gear, this requires two settings of the receiver but the equipment is being modified by the addition of one antenna to complete the set of three. The modified receiver may be mounted solidly at a fixed position. For the basic studies this will eliminate the difficulty involved in keeping the antennas in a vertical line while being raised.

II. THE SOLUTION

The problem[6] is to separate the direct and reflected wave components from a microwave transmitter received at a distance large enough to justify the plane wave approximation. Consider three antennas spaced Δh feet apart vertically and labelled r, s, t, as shown in Fig. 1.

A direct wave of magnitude E and phase angle δ (labeled γ in Fig. 1) is incident on antenna s at an angle of α' above the horizontal. A second wave (the ground reflected component) arrives at an

* This work was performed at The University of Texas under contract N5ori-136, T. O. I. with the Office of Naval Research.

** Now at Reeves Instrument Company, New York, New York.

*** Now at the School of Electrical Engineering, Cornell University, Ithaca, New York.

[1] F. E. Brooks, Jr. and C. W. Tolbert, "Equipment for measuring angle-of-arrival by the phase difference method," The University of Texas, Electrical Engineering Research Laboratory, Report No. 2 (May 16, 1946).

[2] E. W. Hamlin and W. E. Gordon, "Calculated and measured phase difference at 3.2 centimeters wave length," Proc. I. R. E. 36, No. 10, 1218–1223 (1948).

[3] A. W. Straiton and J. R. Gerhardt, "Results of horizontal microwave angle-of-arrival measurements by the phase difference method," Proc. I. R. E. 36, No. 7, 916–922 (1948).

[4] A. W. Straiton, "Vertical phase front measurements for microwave transmission over water near Corpus Christi, Texas," The University of Texas, Electrical Engineering Research Laboratory, Memorandum No. 4 (January 6, 1947).

[5] E. W. Hamlin and A. W. Straiton, "Phase difference between the fields of two vertically spaced antenna," Proc. I. R. E. 36, 1538–1543 (1948).

[6] A. W. Straiton and W. E. Gordon, "A method of determining angle-of-arrival," J. App. Phys. 19, 524–533 (1948).

angle of $\rho' < \alpha'$ with a magnitude of KE and phase of θ at s.

Denote the phase lead of r over s as a result of the angle of the direct wave† by $\alpha = (2\pi/\lambda)\Delta h\alpha'$, and the phase lead of r over s due to the angle of the reflected wave by $\rho = (2\pi/\lambda)\Delta h\rho'$. Let $\Delta\phi_1$, be the absolute phase difference between antennas $(r-s)$, and $\Delta\phi_2$, the phase difference of antennas $(s-t)$.

Assume that only one pair of waves are present, and that α and ρ are constant over the height interval of the three antennas $(2\Delta h)$. The latter assumption is equivalent to a constant reflection coefficient (K) and constant phase lag of the reflected behind the direct component $(\delta-\theta)$, in the same height interval.

The fields at the three antennas may now be expressed using E_s as a reference vector.

$$\begin{aligned} E_r\lfloor\underline{\Delta\phi_1} &= E\lfloor\underline{\delta+\alpha} + KE\lfloor\underline{\theta+\rho}, \\ E_s\lfloor\underline{0} &= E\lfloor\underline{\delta} + KE\lfloor\underline{\theta}, \\ E_t\lfloor\underline{-\Delta\phi_2} &= E\lfloor\underline{\delta-\alpha} + KE\lfloor\underline{\theta-\rho}. \end{aligned} \tag{1}$$

Dividing Eqs. (1) by the magnitude E_s and letting the ratios of signal strength be $R_1 = E_r/E_s$, $R_2 = E_t/E_s$, and $R = E/E_s$;

$$\begin{aligned} R(1\lfloor\underline{\delta+\alpha} + K\lfloor\underline{\theta+\rho}) &= R_1\lfloor\underline{\Delta\phi_1}, \\ R(1\lfloor\underline{\delta} + K\lfloor\underline{\theta}) &= 1\lfloor\underline{0}, \\ R(1\lfloor\underline{\delta-\alpha} + K\lfloor\underline{\theta-\rho}) &= R_2\lfloor\underline{-\Delta\phi_2}. \end{aligned} \tag{2}$$

In Eqs. (2) the known or measured data are R_1, R_2, $\Delta\phi_1$, and $\Delta\phi_2$. There remain six unknowns; R, K, δ, θ, α, and ρ. It should be noted that the value R determines the magnitude of the direct wave component.

The procedure is to equate the real and imaginary parts of each of Eqs. (2), thus obtaining six equations in six unknowns. These will be solved first for the angles α and ρ, which solution is accomplished with good fortune and relative ease.

From (2)

$$R_1\cos\Delta\phi_1 = R[\cos(\delta+\alpha) + K\cos(\theta+\rho)], \tag{3}$$

$$R_1\sin\Delta\phi_1 = R[\sin(\delta+\alpha) + K\sin(\theta+\rho)], \tag{4}$$

$$1 = R[\cos\delta + K\cos\theta], \tag{5}$$

$$0 = R[\sin\delta + K\sin\theta], \tag{6}$$

$$R_2\cos\Delta\phi_2 = R[\cos(\delta-\alpha) + K\cos(\theta-\rho)], \tag{7}$$

$$-R_2\sin\Delta\phi_2 = R[\sin(\delta-\alpha) + K\sin(\theta-\rho)]. \tag{8}$$

† λ is wave-length.

From (3) and (7),

$$\begin{aligned} a &= R_1\cos\Delta\phi_1 + R_2\cos\Delta\phi_2 \\ &= 2R[\cos\delta\cos\alpha + K\cos\theta\cos\rho], \end{aligned} \tag{9}$$

$$\begin{aligned} a_1 &= R_2\cos\Delta\phi_2 - R_1\cos\Delta\phi_1 \\ &= 2R[\sin\delta\sin\alpha + K\sin\theta\sin\rho], \end{aligned} \tag{10}$$

and from (4) and (8),

$$\begin{aligned} b &= R_1\sin\Delta\phi_1 - R_2\sin\Delta\phi_2 \\ &= 2R[\sin\delta\cos\alpha + K\sin\theta\cos\rho], \end{aligned} \tag{11}$$

$$\begin{aligned} b_1 &= R_1\sin\Delta\phi_1 + R_2\sin\Delta\phi_2 \\ &= 2R[\cos\delta\sin\alpha + K\cos\theta\sin\rho]. \end{aligned} \tag{12}$$

Eliminate δ in Eqs. (9) to (12) by use of (5) and (6).

$$a = 2R[(1/R - K\cos\theta)\cos\alpha + K\cos\theta\cos\rho], \tag{13}$$

$$a_1 = 2R[-K\sin\theta\sin\alpha + K\sin\theta\sin\rho], \tag{14}$$

$$b = 2R[-K\sin\theta\cos\alpha + K\sin\theta\cos\rho], \tag{15}$$

$$b_1 = 2R[(1/R - K\cos\theta)\sin\alpha + K\cos\theta\sin\rho]. \tag{16}$$

Solving (13) and (16) for $RK\cos\theta$, equating and simplifying,

$$RK\cos\theta = \frac{\cos\alpha - a/2}{\cos\alpha - \cos\rho} = \frac{\sin\alpha - b_1/2}{\sin\alpha - \sin\rho}. \tag{17}$$

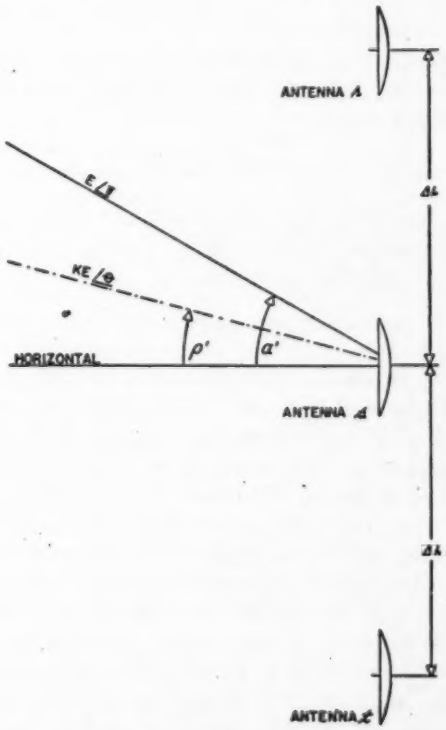


FIG. 1. Notation for solution.

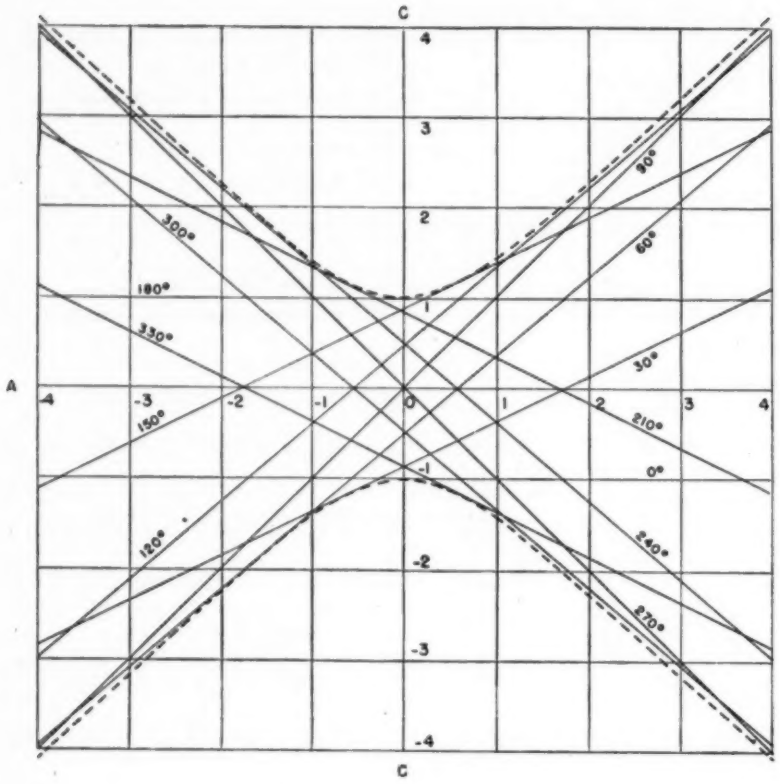


FIG. 2. Evaluation of α or ρ from $A\sin\alpha=\cos\alpha+C$, $A\sin\rho=\cos\rho+C$.

Similarly from (14) and (15),

$$-2RK\sin\theta=\frac{a_1}{\sin\alpha-\sin\rho}=\frac{b}{\cos\alpha-\cos\rho}. \quad (18)$$

To climax the good fortune, divide the right side equations of (17) by (18), and let $A=b/a_1$, and $C=1/2(b_1A-a)$,

$$A\sin\alpha=\cos\alpha+C. \quad (19)$$

Using the procedure employed following Eq. (12), this time eliminating θ, there results an equation similar to (19) but containing ρ.

$$A\sin\rho=\cos\rho+C. \quad (20)$$

It may be noted that Eq. (19) or (20) has two roots between 0 and 360° for a given set of constants A, C (Fig. 2). However, because of the nature of the equation, the value of α (or ρ) will be ambiguous in steps of 360-phase degrees. The magnitude of the geometric angle ambiguity depends on the spacing of the antennas and the wave-length, i.e., $\alpha=(2\pi/\lambda)\Delta h\alpha'$. Approximate knowledge of the geometry and a coarser method of measurement make the selection of the correct angle possible.

The two angles, α, ρ, are of prime importance for the basic angle-of-arrival studies and the possible applications to precise elevation angles. However, the remaining variables, especially R and K, appear to be interesting tools for the study of the mechanism of reflection. Therefore, to complete the solution and find θ in terms of α from (17) and (18);

$$\tan\theta=b/(a-2\cos\alpha), \quad (21)$$

and similarly δ in terms of ρ,

$$\tan\delta=b/(a-2\cos\rho). \quad (22)$$

R and K are then found from (5) and (6).

$$R=\sin\theta/\sin(\theta-\delta), \quad (23)$$

and

$$K=-\sin\delta/\sin\theta. \quad (24)$$

III. TEST OF THE METHOD

The solution presented above has been tested by the evaluation of the angles for sixty sample sets of data taken at the U. S. Naval Electronics Laboratory desert path in Arizona during the spring of 1947.

An example of these calculations is as follows:

Auxiliary Data

Path Length = 27 miles $\lambda=3.2$ cm

Height of Transmitter = 200 feet

Height of Receivers: $t=170$ feet, $s=180$ feet, $r=190$ feet

Necessary Data

$R_1 = .933$

$R_2 = 1.01$

$\Delta\phi_1 = -118°$

$\Delta\phi_2 = -108°$

Calculations

From Eqs. (9), (10), (11), and (12)

$a = -0.750 \qquad b = 0.133$

$a_1 = 0.126 \qquad b_1 = -1.781$

$A = b/a_1 = 1.055$

$C = 1/2(b_1 A - a) = -0.565.$

For these values of A and C, the angle α and ρ are found to be 250° and 20°. But from the geometry the angles-of-arrival are expected to fall between $-100°$ and $-360°$. Since α and ρ are ambiguous in steps of 360°, this angle is subtracted from the chart value to obtain angles of $-110°$ and $-340°$.

The geometric angle-of-arrival, α' and ρ' are found to be

$$\alpha' = -110/600 = -0.12°,$$
$$\rho' = -340/600 = -0.57°.$$

Bending of Beams with Creep

E. P. Popov*

Department of Engineering, University of California, Berkeley, California

(Received July 12, 1948)

A method of calculating stresses and deflections for beams whose material creeps is presented in this paper. Complete tension creep test data at constant temperature are used to define creep characteristics of the material. Then by using Bernoulli's hypothesis of plane sections and the techniques developed earlier for interpretation of the relaxation creep tests, a method of beam analysis is shown. Stresses and deflections may be calculated for any desired time interval. This includes the time prior to the occurrence of the steady state creep. The latter aspect appears to have been ignored by others.

INTRODUCTION

IN some industrial applications materials are used at elevated temperatures. With time the initial elastic stress distribution and deformations change as a result of creep or gradual plastic deformation. The creep phenomenon is likewise clearly exhibited by some important materials such as lead and certain plastics at room temperature. Solutions for stresses and deformations based on the usual elastic theory for such situations are inadequate.

Prediction and interpretation of the behavior of materials with creep considerations attracted several investigators. Most of the attention was properly directed at the steady state of creep. This is logical as it was repeatedly observed that nearly a constant rate of plastic deformation takes place after a relatively short initial interval of time, provided the stress and the temperature remain constant. For the steady state conditions several empirical formulae or "laws" are proposed. Such laws connect the stress with the steady creep rates that would prevail at a constant temperature. These laws are essentially empirical and they completely ignore the transient creep rates. Yet the transient creep in certain cases turns out to be the dominant factor.[1]

When creep is involved, several simple problems of stress distribution and deformation still exist. Here we will confine our attention to pure flexure of a beam at constant temperature. However, the initial or the transient creep rates are considered. This refinement is desirable for a more accurate prediction of beam deflections and becomes of utmost importance in the study of columns whose material creeps. The latter problem appears to be of immediate practical importance.**

Previous investigators in studying bending of prismatic beams with creep either utilized directly or verified experimentally Bernoulli's hypothesis of plane sections. Thus, like in all technical theory of beams, a plane section before bending remains plane. The studies generally were made for the steady state conditions. For the steady creep, for example, the exponential creep law is $u^* = B\sigma^n$, where u^* is the steady creep rate, σ is the stress, and n and B are experimental constants. Then it follows from Bernoulli's assumption that during creep the plastic strains must vary linearly from the neutral axis to maintain a plane section. Written algebraically,

$$\frac{y}{h} = \frac{u_y^*}{u_h^*} = \frac{B\sigma^n}{B\sigma_{\max}{}^n} = \frac{\sigma^n}{\sigma_{\max}{}^n}, \tag{1}$$

where σ is the stress in a fiber at a distance y from the neutral axis, and $\sigma_{\max}$ is the stress in the extreme fiber at $y = h$. For simplicity here we are considering

* Associate Professor of Civil Engineering.

[1] E. P. Popov, J. App. Mech. 14, A-135 (1947).

** A private communication from F. R. Shanley.

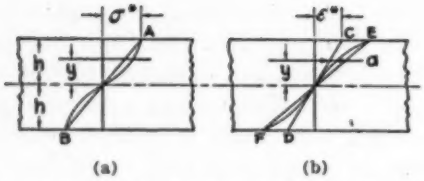


FIG. 1. (a) Stress distribution. (b) Strain distribution.

a rectangular section of a beam. With a further proviso that creep properties of material are alike in tension and compression, the bending moment M for a rectangular beam of width b is

$$M=2\int_0^h \sigma y b dy. \tag{2}$$

This is readily integrated by changing the variables. The expression for stress in any fiber during the steady creep becomes[2,3]

$$\sigma=\left(\frac{y}{h}\right)^{1/n}\frac{Mh}{I}\left(\frac{2n+1}{3n}\right). \tag{3}$$

Here I is the moment of inertia of the cross-sectional area. Other symbols are as defined previously.

Any other creep law may be used to obtain a similar solution. The hyperbolic sine creep law[4] expressing the creep rate as a function of stress is

$$u^*=u_0\sinh\frac{\sigma}{\sigma_0}, \tag{4}$$

where u_0 and σ_0 are experimental constants. The solution for stress distribution in bending in conformity with this law is a mere mathematical exercise, yielding

$$M=\frac{3I}{2h}\left(\sigma_{\max}-\frac{\sigma_0}{2}\coth\frac{\sigma_{\max}}{\sigma_0}+\frac{\sigma_{\max}}{2}\operatorname{csch}^2\frac{\sigma_{\max}}{\sigma_0}\right), \tag{5}$$

and

$$y=h\cdot\sinh\frac{\sigma}{\sigma_0}\Big/\sinh\frac{\sigma_{\max}}{\sigma_0}. \tag{6}$$

These two equations give a complete solution for the steady stress distribution in a rectangular beam. Solution of these equations may be obtained by trial. The form of these expressions is decidedly less convenient than that of Eq. (3), although it is not implied that this should be the main criterion.

Some creep laws that are written[5] so as to depend on time appear to consider the transient characteristic of the creep phenomena. These must not be based on the time-hardening assumption.[1] If used the results obtained in this manner for bending with creep become peculiar. The solution indicates the same stress distribution regardless of the time elapsed. No time appears in the end result. This must be viewed with caution.

Another creep law that is rather general has been used[6] to determine stresses and deflections of beams in bending. Its expression is

$$\sigma=K\epsilon_p^m u^n, \tag{7}$$

where K, m, and n are experimental constants, ϵ_p is the plastic strain, and u is creep rate. In this equation σ, ϵ_p, and u are all variables. It appears to be properly formulated for the transient as well as for the steady creep rates. Then by virtue of Bernoulli's hypothesis

$$\epsilon_y=(y/h)\epsilon_{\max}, \tag{8}$$

where ϵ_y must be the total unit strain (elastic and plastic) at a fiber at a distance y from the neutral axis, and $\epsilon_{\max}$ is the total unit strain in the extreme fibre. Not only plastic but also the elastic strains change with time until the steady state is reached. The method cited ignores the elastic part. Thus the subsequent conclusions seem questionable. Moreover several of the constants involved in Eq. (7) are very difficult to determine experimentally and are generally not available.

STRESS ANALYSIS

Let us consider for simplicity a rectangular prismatic beam in pure flexure. The applied moments act in the plane of symmetry of the cross-sectional area. The material properties including creep characteristics are assumed to be alike in tension and compression. At any section through such a beam the shear is zero. This conveniently omits the necessity of considering shear or variation of moment along the length of the beam. A similar

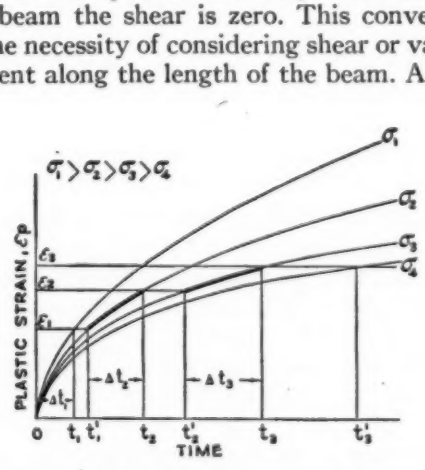


FIG. 2. A family of tension test creep curves at constant temperature.

[2] R. W. Bailey, Proc. Inst. Mech. Eng. 131, 131 (1935).
[3] G. H. McCullough, Trans. A. S. M. E. 55, 55 (1933).
[4] A. Nadai, *Stephen Timoshenko 60th Anniversary Volume* (The Macmillan Company, New York, 1938), p. 155.
[5] C. R. Soderberg, *Stephen Timoshenko 60th Anniversary Volume* (The Macmillan Company, New York, 1938), p. 197.
[6] E. A. Davis, J. App. Mech. 5, A-29 (1938).

simplification is made in the derivation of the flexure formula in the technical elastic theory.

Bernoulli's hypothesis of plane sections is assumed to hold at all times. This notion is fundamental in the procedure adopted. Its validity has been experimentally demonstrated[7] for materials that do not obey Hooke's law and holds above the yield point of the material. In bending with creep the assumption is also valid.[3, 8]

The stress-strain-time relations obtained from the tension creep tests are used for determining the stress distribution and in calculating the deflections. Any expression for the creep strain that satisfactorily fits the tension test data may be used. Mathematical formulation is not basic as long as the expressions adequately describe the entire family of creep curves including the transient portions. The formulae are merely an expedient and may be chosen on the basis of their convenience for calculations.

For a beam in bending the requirements of static equilibrium must be satisfied at all sections and at all times. This may be formulated by the usual two equations: $\sum F=0$, i.e., the summation of all the forces perpendicular to any cross section of the beam must be zero, and $\sum M=M$, i.e., the summation of all of the above internal forces forming the internal resisting moment must be equal to the externally applied moment.

The first requirement is met in the elastic theory by having the neutral plane pass through the centroid of the section. In bending with creep, where stress variation does not follow a simple law, the neutral plane does not necessarily pass through the centroid of a section. However, for cross sections with two axes of symmetry, it does pass through the center of the section, because we have assumed that the properties of materials are the same in tension and compression.

Now let us consider a rectangular beam in pure flexure where material creeps. At the instant that the couples are applied, i.e., $t=0$, the stress distribution and fiber elongation is the same as is normally given by the elastic theory. The material as yet had no time to creep. The stress and strain distribution is instantaneously linear and is represented by the straight lines AB and CD in the respective diagrams of Fig. 1. The corresponding maximum stress is $\sigma^*=Mh/I$ and the maximum strain $\epsilon^*=\sigma^*/E$. Initially stress for any other fibre is $\sigma=(y/h)\sigma^*$. For a small interval of time Δt the fibres are essentially under the action of the above stress distribution. This being the case, we may calculate by any suitable expression for creep the

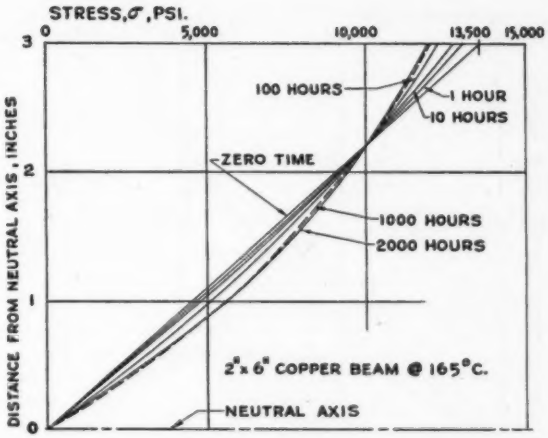


FIG. 3. Stress distribution in a beam.

elongation (or contraction) that will take place in the interval of time $t_1=\Delta t_1$. The fibres farther away will creep more than the ones near the neutral axis as they are under a higher stress. The unit creep elongation added to the initial unit elastic elongations may be represented by the curved line EF in Fig. 1b. But this results in a section that is not plane. So to be consistent with our fundamental assumption, we must increase the stresses in the intermediate fibres by a sufficient amount to make the section plane. From experiments it is known that the material, regardless of the amount of creep that has taken place, has essentially a constant elastic modulus. A fibre at a distance y from the neutral plane must have an elastic increment of stress to make the section plane. This increment must be equal to the difference between the unit strain values of the straight line EF and the curved line EF multiplied by E. Stated in the form of an equation we must have

$$\Delta\sigma_y=((y/h)\widehat{CE}-\alpha)E, \tag{9}$$

where all quantities on the right-hand side of the expression are known, as CE is the plastic strain of the extreme fibre and α is the plastic strain of a fibre y.

These stress increments in our symmetrical case do not disturb the location of the neutral axis, nor $\sum F=0$, but they do the moment. Summation of these forces around the neutral axis gives us an increase in the resisting moment ΔM, but the applied moment is constant. This we compensate by calculating with elastic formulae the adjusted stresses and strains which in turn are added to the quantities just computed. The adjusted stress at a distance y from the neutral axis is

$$\sigma_y'=\frac{(M-\Delta M)y}{I}, \tag{10}$$

[7] A. Nadai, *Plasticity* (McGraw-Hill Book Company, Inc., 1931), p. 160.

[8] H. J. Tapsell and A. E. Johnson, J. Inst. of Metals, **LVII**, pt. 2, 121 (1935).

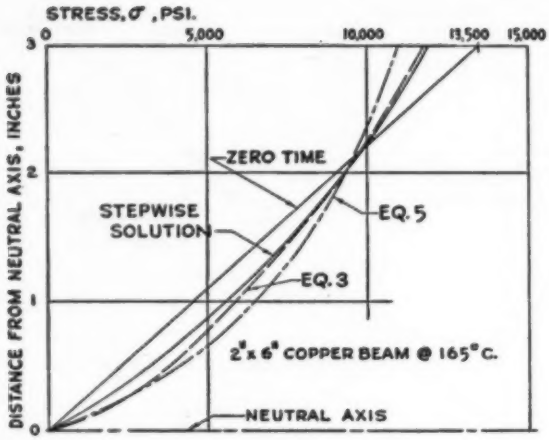


FIG. 4. Steady state stress distribution in a beam.

and the corresponding strain is

$$\epsilon_y' = \sigma_y'/E. \tag{11}$$

This procedure does not change any of the previous reasoning as ϵ' may be used instead of ϵ^* in the strain diagram of Fig. 1. In this manner we arrive at the new time t_1 with all of the requirements of creep, plane section and static equilibrium satisfied. The new stress distribution for any fibre thus becomes

$$\sigma_y = \sigma_y' + \Delta\sigma_y. \tag{12}$$

To determine stress distribution for a longer period of time we proceed in a similar way. A new Δt_2 is assumed. The new starting stress distribution is now the one just computed. Here we may note that the behaviour of the individual fibres of a beam subjected to uni-axial stress is similar to the case of a stressed creeping bolt that clamps rigid flanges. This problem was discussed by the writer previously,[1] and good agreement was found to exist between the tension creep tests and the relaxation tests. The same technique of analysis no doubt may be satisfactorily applied here.

Good solutions for fibres in relaxation were obtained on the basis of the strain-hardening assumption. This principle is illustrated in Fig. 2. In this figure we first proceed along the curve σ_1 of constant stress for a time interval t_1 to obtain the first amount of the plastic strain ϵ_1. Then suppose we find that the stress for the particular fibre in the interval of time t_1 has changed to a new stress σ_2 from Eq. (12). For the next selected time interval Δt_2 material creeps essentially under stress σ_2 and according to the strain-hardening assumption now we start not from the time t_1 but from the time t_1' and proceed to the time $t_1' + \Delta t_2$. After the new time interval we arrive at plastic strain ϵ_2. In this manner a new plastic strain of several fibres may be computed. Repeating the adjustment procedure for the actual moment, the stress distribution at the end of the time $t_1 + \Delta t_2$ is obtained. Stress distribution at any time may be calculated by making additional steps.

A little further reflection on the behavior of the beam's fibres shows some very interesting things. With time the initially highly stressed fibres become somewhat relieved of the stress. The interior ones on the other hand contribute gradually more and more to the resisting moment. Eventually the state of stress distribution is reached when the creep rates in various fibres become proportional to their respective distances from the neutral plane. Any other situation is transient in character. This is the steady state—for it, as mentioned earlier, satisfactory solutions are available.[2,3,8] The solution presented is in complete accord with such investigations.

Obviously the time necessary to arrive at the steady state stress distribution depends on the material's creep characteristics and will be different for various materials. The stable stress distribution occurs and continues to exist when the creep rates in the fibres are proportional to their respective distances from the neutral axis. In any beam that exhibits creep such a condition eventually is attained. And if there is sufficient time for the creep in all fibres to pass over the transient period and to attain the minimum creep rates, a direct solution using a minimum creep law is possible. The creep rates eventually become minimum and their variation with stress is well described by the mathematically expressed laws. Remarkably, the previous stress distribution is of no importance. This fact gives the reason for the good experimental agreement found under the steady state conditions with formulae based on the minimum creep laws.

NUMERICAL EXAMPLE

Let us assume a rectangular beam that is 2-in. wide by 6-in. deep subjected to end moments of 162,000 in.-lb. This beam is to be of oxygen-free copper with the elastic modulus $E = 14.1 \times 10^6$ p.s.i. This material has been carefully investigated[9] and its tension test creep curves at 165°C may be approximated[1] by the following expression

$$\epsilon_p = 26.8 \times 10^{-6}(e^{\sigma/7500} - 1)t^{0.372} \text{ in./in.} \tag{13}$$

This equation is of a parabolic type. Thus no

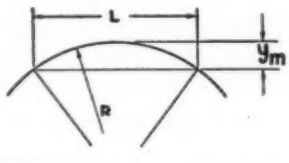


FIG. 5. Elastic curve of a deflected beam (exaggerated).

[9] E. A. Davis, J. App. Mech. 10, A-101 (1943).

minimum creep rate is ever reached, but it has been found[1] to be satisfactory to nearly 5000 hours as determined by verification with the relaxation test.

With this data calculations were carried out for 1, 3, 10, 50, 100, 200, 400, 1000, and 2000 hours. The time intervals were gradually increased as the creep curves flatten out and the fibre stresses became more stabilized. For calculations the beam was subdivided vertically into one-half inch intervals from the neutral axis. Initially elastic distribution was assumed and calculations were carried forward as explained previously. Simpson's rule was used to obtain ΔM. The results of the calculations are shown in Fig. 3. Reduction in the extreme fibre stresses is here clearly seen.

The copper considered apparently is sluggish in its attainment of the stabilized stress distribution. Much more rapid stress redistribution is reported by some investigators for lead.[8] The reason for this becomes apparent by considering the exponential form of the creep law. For the copper considered we have $u^* = B\sigma^{1.6}$ and for lead $u^* = B\sigma^{5.8}$. The exponent of the stress in these formulae may be thought of as an index of the materials' sensitivity to creep as affected by the stress level. For the copper beam considered, the sensitivity to stress variation is far less than would be expected for a material such as lead. Moreover the transient part of the tension creep test for lead is of very short duration.

The steady stress distribution as calculated is further compared with solutions from Eqs. (3), (5), and (6) in Fig. 4. Fair agreement is found with the exponential creep law, the one that is experimentally verified,[8] and poor agreement is found if the hyperbolic sine creep law is used. One could speculate on this point and wonder whether the undesirable characteristic of the hyperbolic sine law is not brought out by this analysis. This law would appear to give excessive creep rates for low stresses. The "conservative" nature of this function for low stresses appears to be indicated.

DEFLECTION OF BEAMS

Here we will consider the deflection of a beam exhibiting creep as a function of time. The same assumptions and limitations as were earlier imposed in establishing the procedure to determine the stress distribution at various times apply here. The problem being considered is for a rectangular beam in pure bending.

From the hypothesis of plane sections, it is known that the following relation holds true

$$1/R = \epsilon_m/h, \tag{14}$$

where R is the radius of curvature of the bent beam,

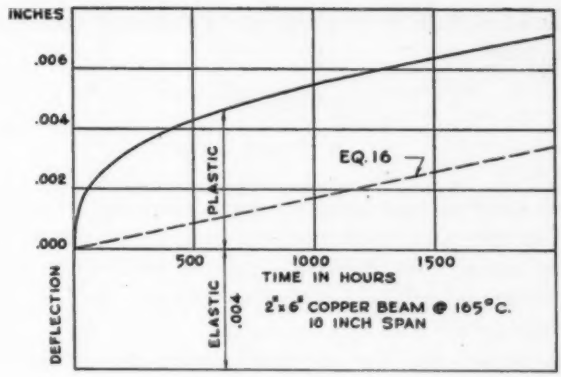


FIG. 6. Beam deflection.

ϵ_m is the unit strain in the extreme fibres, and h is one-half depth of the beam.

In considering the deflection of beams the actual deflections are small, and for pure flexure, the radius of the elastic curve is a constant. Recognizing the smallness of the deflection investigated it can be expressed with sufficient accuracy by one term of a series. Thus,

$$y_m = (L^2\epsilon_m/8h) \tag{15}$$

where y_m is the deflection at the middle of a beam between two gage points L inches apart, and as before h is half of the depth of the beam, and ϵ_m is the unit strain (elastic and plastic) in the extreme fibres. The meaning of the terms may become clearer by referring to Fig. 5.

When we considered earlier the stresses in a beam as a function of time the elastic and plastic strains in the extreme fibres were known. Their sum is the quantity ϵ_m in Eq. (15). Further, if we know the distance between certain gage points or between the points of application of the couples, the value of L is known. Thus, for pure flexure, deflection as a function of time immediately follows.

Considering the hypothetical beam previously analyzed and assuming $L = 10$ in., calculations were performed to obtain the deflection at the center of the beam as a function of time. The initial deflection is entirely elastic. Subsequent increments are caused by the change of strains in the extreme fibres. The difference between the new and the preceding creep strain added algebraically to the change in the elastic strain is ϵ_m.

A plot of the center deflection as a function of time for the assumed beam is shown in Fig. 6. On the same plot a line is shown which is calculated from equations which are based on the minimum creep law of the exponential type. These equations are[10]

$$y_m = (t/D)(L^2/8)M^n, \tag{16}$$

[10] J. Marin and L. E. Zwissler, Proc. A. S. T. M. 40, 937 (1940).

where t is time, n is the stress exponent in the creep law, and

$$D=\frac{1}{B}\frac{(2b)^n h^{2n+1}}{[2+(1/n)]^n} \tag{17}$$

with symbols that have the same meaning as before.

For the particular material considered the transient creep characteristic appears important. For very long intervals of time Eq. (16) may be entirely satisfactory.

A replot for the assumed beam on a log-log system of coordinate axes of creep deflection at center *versus* time was found to lie remarkably close to a sloping straight line. This fact is highly significant. All of our reasoning was based on the creep data derived from a tension creep test. Precisely the same thing is found experimentally. The results of an extensive series of tests[11] on aluminum alloys, copper and monel yielded such graphs. This appears to be the usual characteristic for uniformly stressed cantilever beams. It has even been suggested[11] to call this slope from a log-log plot the "creep susceptibility factor." No tie-in existed between these results and the tension creep test. The foregoing offers an explanation.

CONCLUSIONS

A step-by-step method of calculations has been presented that gives stress distribution and deflections of a beam in pure bending at any time. The solution is based on data derived from the tension creep test. The effect of the transient period of creep on deflections may be marked. This may be of great importance in the study of excentrically loaded columns that creep.

For beams that are slender and not in pure bending the shear deflection can probably be neglected. Such beams may be subdivided into a number of sections. Each section then can be treated as if it were in pure bending: the deflections of each section computed, and the total calculated by "piecing together" all of the subdivisions.

Published results of experiments by others completely substantiate the suggested procedures. Answers in the foregoing discussion were offered to some of the disconnected observations of creep testing. An attempt to unify them was made.

Naturally all of the discussion was confined to the primary and the secondary creep stages. The tertiary zone was not considered. Metallurgical considerations are ruled out, only materials found suitable for the particular temperature may be analyzed by the theory presented. Methods of calculations in the form presented are valid only for situations at constant temperature.

At present, need still exists for an all inclusive creep law. Such a law should properly describe the creep deformation in the transient as well as in the steady creep zones. It also should show the dependence of creep rates on temperature. Attempts in this direction have been made. Such laws as yet appear to be specialized. Until the time when we have a general creep law we may have to continue the use of arithmetical methods, laborious as they may be.

ACKNOWLEDGMENTS

The writer acknowledges with gratitude the advice received during the progress of this work from Professor S. Timoshenko, under whose general direction the study was made. The above discussion mainly presents an abridged chapter from a doctoral dissertation presented at Stanford University. Other portions were published earlier.[1,12]

[11] R. G. Sturm, C. Dumont, and F. M. Howell, J. App. Mech. **3**, A-62 (1936).

[12] E. P. Popov, J. Franklin Inst. **243**, 365 (1947).

Analysis of the Metal-Strip Delay Structure for Microwave Lenses

S. B. Cohn

Radio Engineering, Sperry Gyroscope Company, Division of the Sperry Corporation, Great Neck, New York

(Received August 3, 1948)

The metal-strip delay medium is shown to be analogous to a transmission-line low pass filter. A filter analysis gives the image phase constant and image admittance of the equivalent filter section. The index of refraction of the medium is obtained from the former and the reflection coefficient from the latter. It is shown that the physical parameters may be chosen so that the strip medium is perfectly matched to space at a particular frequency. Formulas are given which hold for normal incidence even when the strips are close together. The same method may be applied to delay media consisting of thin metal obstacles of any shape arranged uniformly in parallel planes. Examples are circular disks, ellipses, and rectangles. Certain general conclusions are drawn for this class of media.

THE structure shown in Fig. 1(a) has proved useful in the construction of lenses for microwaves.[1] Figure 1(b) shows how such a structure may be shaped so as to focus energy from a point source into a sharp beam. Kock[1] suggests practical methods of construction and gives the lens-shape formula.

THE ORIGINAL ANALYSIS

Kock arrived at this and other dispersed metal structures through an analogy to light transmission in refracting media. Each isolated strip is electrically polarized by an electromagnetic radio wave in an analogous manner to the way molecules in a transparent body are polarized by an electromagnetic lightwave. The strips and the molecules both, therefore, increase the dielectric constant of space. The index of refraction, which is the ratio of phase velocity in space to that in the medium, is related to dielectric constant by

$$\eta = (\kappa)^{\frac{1}{2}}, \tag{1}$$

where η is the index of refraction and κ the relative dielectric constant of the medium. At low frequencies and for strips narrow with respect to their spacings, the dielectric constant of the strip lens is given by Kock as

$$\kappa = 1 + (\pi/4)(w^2/bl), \tag{2}$$

where w, b, and l are dimensions defined by Fig. 1.

Kock takes the change in dielectric constant with frequency into effect by means of the following approximate relation:

$$\kappa' = 1 + \frac{\kappa - 1}{1 - (2w/\lambda)^2}, \tag{3}$$

where κ' is the dielectric constant for wave-length λ and κ is the static value for $\lambda = \infty$.

[1] W. E. Kock, "Metallic delay lenses," Bell. Sys. Tech. J. 27, 58 (1948).

Equations (2) and (3) are exact only if the strips are not too closely spaced and if the wave-length is not too small. For the close spacings used in practical lenses they are not of sufficient accuracy to eliminate cut and try in the design of the lens and therefore a more complete formula for index of refraction is desirable. Also a formula giving the reflection coefficient at the surface of a strip medium would be useful.

TRANSMISSION-LINE FILTER APPROACH

If normal incidence to a plane strip-lens surface is assumed, the dotted planes shown in Fig. 2(a) will be everywhere perpendicular to the electric field and parallel to the magnetic field. Thin conducting sheets may therefore be placed in the dotted planes without altering the electromagnetic field. The portion of the lens thus isolated is shown in Fig. 2(b), where a constant width a in the x direction is assumed. The discontinuity resulting from limiting the x-direction width in this manner may be neglected if $a \gg b$.

Figure 2(b) is seen to represent a parallel-strip transmission line loaded by a succession of shunt capacitive irises.* The characteristic impedance of this transmission line is equal to that of free space; that is, if a were equal to b, the characteristic impedance would be 120π-ohms, which is the characteristic impedance of free space per square. In the following work, all impedances are normalized with respect to the transmission-line characteristic impedance, and hence the impedance of free space is taken as unity.

The equivalent circuit of Fig. 2(b) is shown in Fig. 3(a). The blocks represent lengths l of transmission line having $Z_0 = 1$. This circuit may be regarded as a filter; one section of which is shown in Fig. 3(b). From the solution of the single section, one may obtain all the properties of the cascade of

* The similarity of the delay structures to capacitively loaded transmission lines was pointed out in Kock's paper.

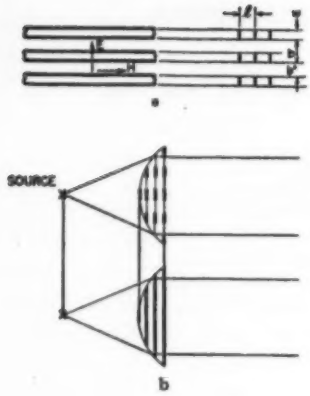


FIG. 1. (a) Delay structure; (b) lens.

sections of Fig. 3(a), and hence all the properties of the strip lens structure of Fig. 2(a) for normal incidence. Note that when a number of sections are connected in cascade, the first and last irises must produce half the susceptance of the inner irises.

Figure 3(b) represents a transmission-line low pass filter section which has been previously analyzed.[2] The image parameters of a single section as given by *Very High Frequency Techniques* are, after simplification,

$$Y_I = (1 + B\cot\varphi - (B^2/4))^{\frac{1}{2}}, \tag{4}$$

$$\beta = \cos^{-1}\{\cos\varphi - (B/2)\sin\varphi\}. \tag{5}$$

Y_I is the image admittance and β the image phase constant of the filter section. (Y_I is the reciprocal of image impedance Z_I.) B is the normalized shunt susceptance of a capacitive iris, and φ is phase length defined by

$$\varphi = 2\pi l/\lambda \text{ radians} = 360l/\lambda \text{ degrees}. \tag{6}$$

INDEX OF REFRACTION FOR $l > b$

The index of refraction is

$$\eta = \beta/\varphi, \tag{7}$$

or

$$\eta = (1/\varphi)\cos^{-1}\{\cos\varphi - (B/2)\sin\varphi\}. \tag{8}$$

The normalized shunt susceptance B of a *thin, isolated* capacitive iris for b/λ small is[3] (Fig. 3c)

$$B = (4b/\lambda)\log_e \csc\pi b'/2b. \tag{9}$$

The irises may be considered isolated if $b < l$.

For the limiting case of $\lambda \to \infty$ and $w/b \to 0$, Eqs. (8) and (9) agree identically with Eqs. (1) and (2). This may be shown through the use of the following limit relationships:

$$\cos^{-1}(\cos\varphi - (B/2)\sin\varphi) \to \cos^{-1}(1 - \varphi^2/2 - B\phi/2) \to (\varphi^2 + B\varphi)^{\frac{1}{2}}$$

and

$$\log_e \csc\pi b'/2b = \log_e \csc(\pi/2 - \pi w/2b) = \log_e \sec\pi w/2b \to \log_e[1 + \tfrac{1}{2}(\pi w/2b)^2] \to \tfrac{1}{2}(\pi w/2b)^2.$$

Equation (9) holds fairly well for b/λ up to 0.2 or so. For b/λ greater than 0.2 the exact formula for a thin, isolated capacitive iris should be used. This is given in the *Waveguide Handbook* and is also represented graphically.[4]

The index of refraction and image admittance are sketched in Fig. 4. As explained in *Very High Frequency Techniques*,[2] filter pass bands occur for Y_I real, and stop bands for Y_I imaginary. The former case corresponds to a transparent lens medium, and the latter to an opaque medium. Additional pass bands occur at higher frequencies than are shown in Fig. 4, but they are of little interest in lens design.

CONDITION FOR NO REFLECTION, $l > b$

The reflection coefficient for a wave entering the lens medium is

$$r = \frac{1 - Y_I}{1 + Y_I}. \tag{10}$$

For an emerging wave, it is minus this. The combined reflection coefficient of a medium of finite thickness may be greater or less than this depending on the phase length between the front and back surfaces. The insertion loss and input standing-wave ratio of the medium as a function of Y_I and the phase length may be obtained from Figs. 26–9 of *Very High Frequency Techniques*.[2] The reflection coefficient is related to the voltage standing-wave ratio S by $|r| = (S-1)/(S+1)$.

Figure 4 shows that a frequency f_0 exists at which Y_I of the lens medium is unity. By Eq. (10), the reflection coefficient is zero for this condition, and there will be no reflection at normal incidence.

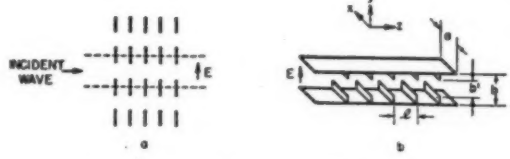


FIG. 2. Equivalent transmission line.

[2] Radio Research Laboratory Staff, *Very High Frequency Techniques* (McGraw-Hill Book Company, Inc., New York, 1947), pp. 659–660.

[3] *Microwave Transmission Design Data* (Sperry Publication No. 23–80, 1944), p. 105.

[4] MIT Radiation Laboratory, *Waveguide Handbook*, Report 41, 1/23/45, Section 1 and Fig. 1-I. Also Radiation Laboratory Series, Volume 10, McGraw-Hill Book Company, New York.

If Y_I is set equal to one in Eq. (6), the following condition results:

$$B_0 = 4\cot\varphi_0. \tag{11}$$

The subscript "0" denotes values at f_0. If Eq. (11) is substituted in Eq. (5), one obtains

$$\beta_0 = \cos^{-1}\left[\cos\varphi_0 - \frac{4\cot\varphi_0}{2}\sin\varphi_0\right]$$

$$= \cos^{-1}(-\cos\varphi_0)$$

$$\beta_0 = \pi - \varphi_0 \text{ radians} = 180 - \varphi_0 \text{ degrees}. \tag{12}$$

Hence,

$$\eta_0 = (180/\varphi_0) - 1, \tag{13}$$

where φ_0 is in degrees. Equation (13) is plotted in Fig. 5.

DISPERSION

Equation (8) indicates that η is a function of frequency. In order to ascertain the width of the frequency band near the operating point f_0 for which the change in η is small, the dispersion of the index of refraction will be calculated. Define this quantity as

$$D = f d\eta/df = \varphi d\eta/d\varphi. \tag{14}$$

Therefore, the change in η is equal to D times the fractional change in f. In the appendix this is found to be

$$D = 2\csc^2\varphi_0 + \frac{2\cot\varphi_0 - \pi}{\varphi_0}. \tag{15}$$

Equation (15) is plotted in Fig. 6. As an example, for $\varphi_0 = 75°$, $D = 0.153$, and over a 10 percent frequency band, the change in η is not more than 0.0153.

The above analysis requires the strips at the front and back surfaces of the lens to have half the susceptance of the inner ones. An edge view of the delay medium will therefore be as shown in Fig. 7.

CASE OF $b > l$

Figure 3 gives the equivalent circuit of Fig. 2 only if the strips are sufficiently far apart horizontally so that the field disturbance set up by each does not extend to the adjacent strips. If $b < l$, the strips may be considered isolated in the horizontal direction. In

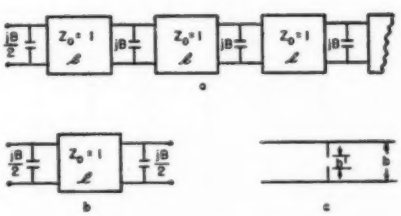


FIG. 3. Equivalent circuit.

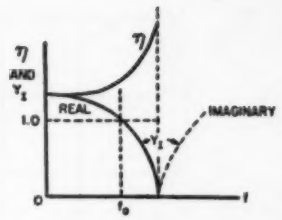


FIG. 4. Index of refraction and image admittance of a strip lens.

a practical lens, however, it is often desirable to have $b > l$ in order to increase the vertical separation of the strips and reduce the number of strips in the lens. The writer[5] has shown previously that for $b > l$ the interaction of the field disturbances has the effect of decreasing the shunt susceptances for given b/λ and b'/b and of introducing bridging capacitive susceptances as shown in Fig. 8. Figure 8 assumes direct interaction only between adjacent strips. If interaction between strips spaced nl apart occurs $(n = 2, 3, 4, \cdots)$, the equivalent circuit will be even more complicated. This effect will undoubtedly be slight in practical lens structures and may be neglected. The filter-section parameters and the index of refraction may be expressed in the following form

$$Y_I = (-B_{oc}B_{sc})^{\frac{1}{2}}, \tag{16}$$

$$\tan\beta/2 = (-B_{oc}/B_{sc})^{\frac{1}{2}}, \tag{17}$$

$$\eta = (\lambda/\pi l)\tan^{-1}(-B_{oc}/B_{sc})^{\frac{1}{2}}. \tag{18}$$

B_{oc} and B_{sc} are, respectively, the normalized open- and short-circuit susceptances of a half-section. These have been derived for the wave-guide filter structure of Fig. 9(a) for $l' \geq b'$. At the plane $z = -l/2$ (Fig. 9b), they are given by

$$B_{oc} = \tan\frac{\pi l}{\lambda} + \frac{2b}{\lambda}\left\{\frac{S_0(\delta)}{\pi^2} - 0.045\right\}$$

$$+ \frac{2b}{\lambda}\sum_{n=1}^{\infty}\left\{\frac{\tanh(n\pi l/b)F}{F} - 1\right\}\frac{\sin^2\pi n\delta}{n(\pi n\delta)^2} \tag{19}$$

$$B_{sc} = -\cot\frac{\pi l}{\lambda} + \frac{2b}{\lambda}\left\{\frac{S_0(\delta)}{\pi^2} - 0.045\right\}$$

$$+ \frac{2b}{\lambda}\sum_{n=1}^{\infty}\left\{\frac{\coth(n\pi l/b)F}{F} - 1\right\}\frac{\sin^2\pi n\delta}{n(\pi n\delta)^2} \tag{20}$$

where

$$\delta = b'/b, \tag{21}$$

$$F = (1 - (b/n\lambda)^2)^{\frac{1}{2}}, \tag{22}$$

[5] S. B. Cohn, *Study of a Wave Guide Filter Structure* (Ph.D. Thesis, Department of Engineering Sciences and Applied Physics, Harvard University, 1948. Also Technical Report No. 39, Cruft Laboratory, Harvard University, April 25, 1948). Also accepted for publication by Proc. I.R.E.

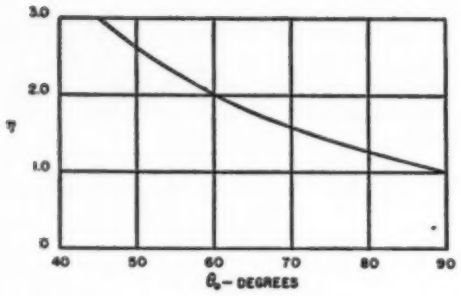


FIG. 5. Index of refraction *vs.* spacing for an image admittance of unity ($l>b$).

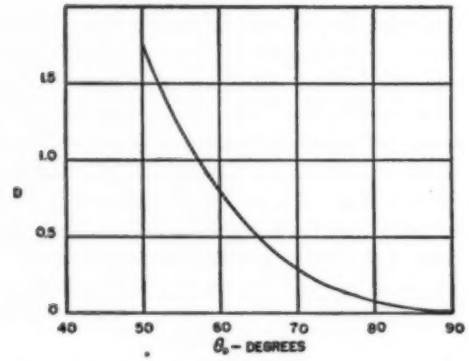


FIG. 6. Dispersion *vs.* spacing for an image admittance of unity ($l>b$).

and where

$$S_0(\delta)=\sum_{n=1}^{\infty}\frac{\sin^2\pi n\delta}{n(\pi n\delta)^2} \tag{23}$$

is the Hahn function of zero order. $S_0(\delta)$ is tabulated by Whinnery and Jamieson.[6] Equations (19) and (20) are valid for all l/b and for b/λ up to 0.9 or so.

In Eq. (19) the first term $\tan(\pi l/\lambda)$ is the input susceptance of the open-circuited line of length $l/2$, the second term is the low frequency value of the step discontinuity susceptance at $z=-l/2$, and the last term takes into account l/b and b/λ. Similar remarks apply to the terms of Eq. (20).

Equations (19) and (20) do not strictly apply to the case of a thin capacitive iris since the discontinuity field of an iris is slightly different from that of a step discontinuity. For the isolated iris of Fig. 3(c), however, the total shunt susceptance is very nearly equal to twice the susceptance of the single step in Fig. 9(b), if b'/b is small. It is also reasonable to expect the dependence on b/λ and l/b of Figs. 2(b) and 9(a) to be similar. For these reasons, Eqs. (19) and (20) should apply fairly well to the thin-strip structure. To improve the accuracy, the second terms of these equations are replaced by half the low frequency susceptance of the iris (Eq. (9)):

$$B_{oc}=\tan\frac{\pi l}{\lambda}+\frac{2b}{\lambda}\log_e\csc\frac{\pi\delta}{2}+\frac{2b}{\lambda}\sum_{n=1}^{\infty}\left[\frac{\tanh(n\pi l/b)F}{F}-1\right]\frac{\sin^2\pi n\delta}{n(\pi n\delta)^2}, \tag{24}$$

$$B_{sc}=-\cot\frac{\pi l}{\lambda}+\frac{2b}{\lambda}\log_e\csc\frac{\pi\delta}{2}+\frac{2b}{\lambda}\sum_{n=1}^{\infty}\left[\frac{\coth(n\pi l/b)F}{F}-1\right]\frac{\sin^2\pi n\delta}{n(\pi n\delta)^2}. \tag{25}$$

Equations (16) and (18) in conjunction with (24) and (25) give the image admittance and index-of-refraction functions of a thin-strip delay lens. These functions are plotted in Figs. 10, 11, 12, and 13 for various values of the physical parameters. Note that a wide variety of parameters is available for which Y_I is equal to or near unity. Since b/λ is proportional to frequency, these curves show the variation of η and Y_I with frequency.

The question of proper strip width at the surfaces of the lens (Fig. 7) is more complicated than before, since a reduction in strip width affects the bridging susceptances of Fig. 8 as well as the shunt susceptances. As an approximation, however, the spacing b' at the surfaces may be increased to the value which reduces Eq. (9) in half. The reflection which results will be small compared to that if the surface strips were not changed at all.

If the strips are embedded in a material of dielectric constant κ which is shaped to the same contour as the strip lens, the index of refraction and Y_I are approximately multiplied by $(\kappa)^{\frac{1}{2}}$. The precise values may be calculated from Eqs. (16) and (18) if in Eqs. (24) and (25) the free-space wave-length λ is replaced by $\lambda/(\kappa)^{\frac{1}{2}}$, and the right sides are multiplied throughout by $(\kappa)^{\frac{1}{2}}$. For polyfoam which has κ approximately equal to 1.02, the increase is about one percent. If, however, the strips are embedded in a

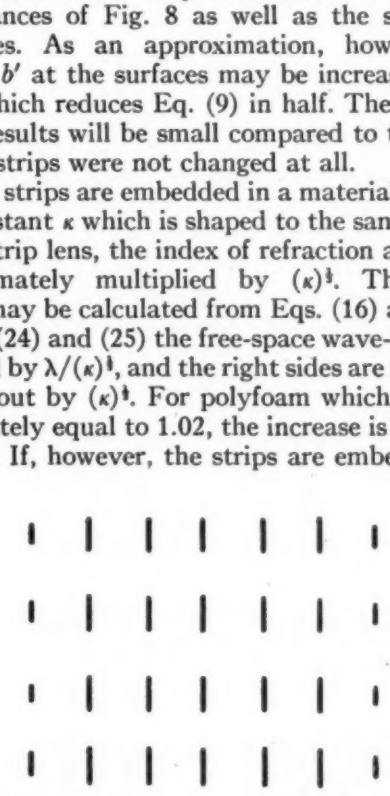


FIG. 7. Change in width of the strips at the surfaces of the medium.

[6] Whinnery and Jamieson, *Equivalent Circuits for Discontinuities in Transmission Lines*, Proc. I.R.E. **32**, 98 (1944).

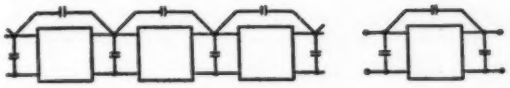


FIG. 8. Equivalent circuit for $b/l>1$.

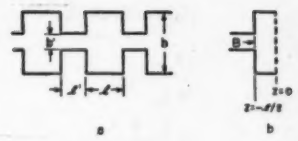


FIG. 9. The wave-guide filter for which B_{0c} and B_{∞} were derived.

rectangular block of polyfoam which has plane-parallel faces instead of the lens shape (as in reference 1), the lens will focus as though it has an index of refraction somewhere between 0 and 1 percent higher than the lens in air.

As an example, consider the strip-medium dimensions given by Kock:[1] $w=0.75$ in., $b=1.3125$ in., $l=0.375$ in. At about 4100 Mc, Kock gives the measured index of refraction as 1.5. For these parameters, values calculated from Eqs. (16) and (18) are $Y_I=1.073$ and $\eta=1.476$. Note that Y_I is near one, and hence the reflections will be fairly low if the surface strips are properly altered.

APPLICATION TO OTHER DELAY STRUCTURES

Although the above study applies specifically to the metal-strip structure, it is also applicable to any delay medium consisting of thin metal obstacles placed uniformly in parallel planes. Examples are arrangements of thin circular disks, ellipses, or rectangles. In such media, an equivalent transmission line may be set up as in Fig. 2(b). If far enough apart horizontally (roughly $l/b>1$), the thin metal obstacles behave as simple shunt susceptances, no matter what their shape may be, and hence the equivalent circuit of Fig. 3 and the entire development of Eqs. (4) to (15) apply exactly. (If the obstacles are not separated sufficiently horizontally, their local field disturbances will not be isolated, and the problem will be more complex as in Fig. 8.) All that is required to complete the theoretical treatment of any of these structures is a determination of the shunt susceptance of the equivalent iris. Even without this quantity, however, the general conclusion may be drawn that any such media will have η and Y_I functions like those of Fig. 4. A frequency can always be found, therefore, for which $Y_I=1$, and hence for which reflections are zero when the surfaces are properly altered.

EXPERIMENTAL TESTS

A cylindrical metal-strip delay lens which provides focusing in the H plane was constructed and tested. The lens dimensions are:

Aperture	25×25 inches
Focal length	20 inches
b	1.125 inches
l	0.375 inch
w of internal strips	0.643 inch
w of surface strips	0.567 inch

At 5000 Mc the calculated index of refraction is 1.48 and the image admittance 1.000. The radiation pattern was measured with a horn feed at the focus, and is as expected for a lens of that aperture.

The reflection from the lens is of the most interest. A slotted line was inserted between the signal source and the horn, and the standing-wave ratio due to the lens was measured on the slotted line with the horn spaced about two inches from the front strips on the curved side of the delay lens. These front strips are $5\frac{1}{2}$ inches long, which is approximately equal to the aperture of the horn, and hence the horn viewed a specimen of essentially uniform thickness. The measured values were corrected so as to eliminate the effect of the reflections from the horn alone. Due to the curvature of the phase front, the first surface of the delay specimen produces a greater effect in the slotted line than the second. This also was taken into account, and the resulting values are plotted in Fig. 14.

The same thickness of a solid dielectric having the same index of refraction would have peaks of $1.48^2=2.19$ for a thickness of $180n+90$ degrees and

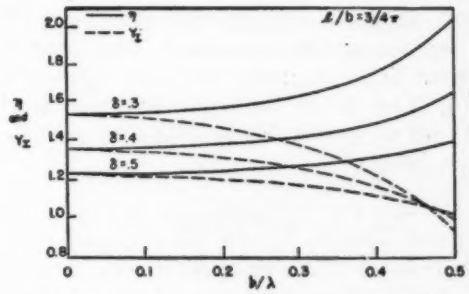


FIG. 10. Index of refraction and image admittance for $l/b=3/4\pi$.

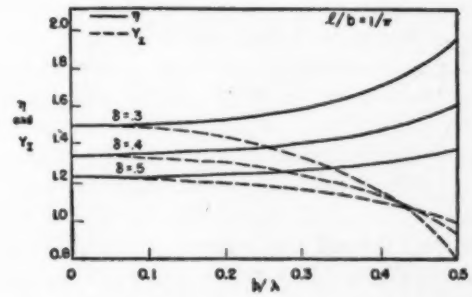


FIG. 11. Index of refraction and image admittance for $l/b=1/\pi$.

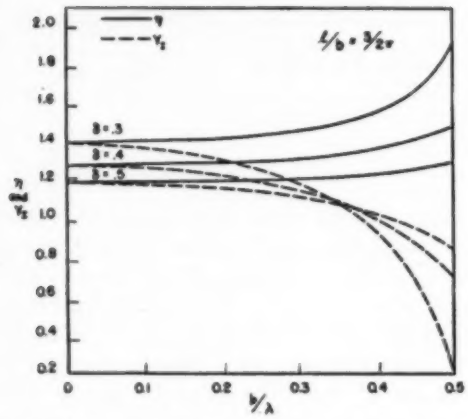


FIG. 12. Index of refraction and image admittance for $l/b=3/2\pi$.

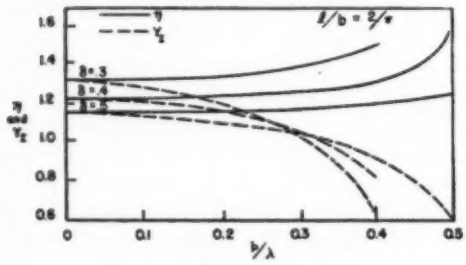


FIG. 13. Index of refraction and image admittance for $l/b=2/\pi$.

minimums for $180n$ degrees. The same thickness of a lens material having $Y_I=1$, but without the modified end strips, would have maxima of 2.66 near $180n$ degrees and minima near $180n+90$ degrees. The theoretical maximum for each case is shown in Fig. 14. The great improvement due to having $Y_I=1$ and to modifying properly the surface strips is clearly evident.

APPENDIX

$$D=f(d\eta/df)=\varphi(d\eta/d\varphi).$$

By Eq. (7)

$$\frac{d\eta}{d\varphi}=\frac{1}{\varphi}\frac{d\beta}{d\varphi}-\frac{\beta}{\varphi^2}=\frac{1}{\varphi}\left(\frac{d\beta}{d\varphi}-\eta\right)$$

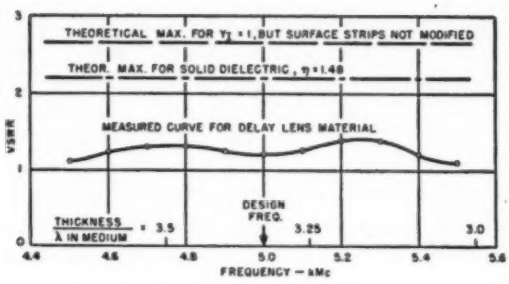

THEORETICAL MAX. FOR Y_I = 1, BUT SURFACE STRIPS NOT MODIFIED
THEOR. MAX. FOR SOLID DIELECTRIC, η = 1.48
MEASURED CURVE FOR DELAY LENS MATERIAL
VSWR
DESIGN FREQ.
THICKNESS / λ IN MEDIUM = 3.5
3.25
3.0
FREQUENCY — kMc

FIG. 14. VSWR of lens material 5.25 inches thick.

and by Eq. (13)

$$\left.\frac{d\beta}{d\varphi}\right|_{\varphi=\varphi_0}-\eta_0=\left.\frac{d\beta}{d\varphi}\right|_{\varphi=\varphi_0}+1-\frac{\pi}{\varphi_0},$$

where φ_0 is in radians.

Since B is closely proportional to frequency and hence to φ, Eq. (5) may be written

$$\beta=\cos^{-1}\left[\cos\varphi-\frac{\varphi}{\varphi_0}\frac{B_0}{2}\sin\varphi\right],$$

$$\frac{d\beta}{d\varphi}=\frac{\sin\varphi+\dfrac{B_0\sin\varphi}{2\varphi_0}+\dfrac{\varphi}{\varphi_0}\dfrac{B_0}{2}\cos\varphi}{(1-(\cos\varphi-(B/2)\sin\varphi)^2)^{\frac{1}{2}}}.$$

For $\varphi=\varphi_0$

$$\left.\frac{d\beta}{d\varphi}\right|_{\varphi=\varphi_0}=\frac{\sin\varphi_0+\dfrac{B_0}{2}\left(\dfrac{\sin\varphi}{\varphi_0}+\cos\varphi\right)}{(1-(\cos\varphi_0-(B_0/2)\sin\varphi_0)^2)^{\frac{1}{2}}}$$

$$=1+\frac{2\cot\varphi_0}{\varphi_0}+2\cot^2\varphi_0.$$

Hence, Eq. (15) results.

The Response of Certain Inorganic Phosphors to Far Ultraviolet Radiation

D. H. Tomboulian and E. M. Pell
Cornell University, Ithaca, New York

(Received August 24, 1948)

The response characteristics of a number of inorganic phosphors have been investigated in the extreme ultraviolet region (~170A) for the purpose of selecting a luminescent screen suitable for the measurement of relative intensities in conjunction with photo-multipliers. Preliminary tests indicate that in this spectral region luminescence is excited rather strongly in the tungstates. Sample screens of $CaWO_4$ have been studied photographically in order to determine to what extent such screens give rise to a faithful reproduction of the intensity distribution of the exciting radiation. Qualitative calculations show that the luminescent intensity excited by weak soft x-ray radiation is above the noise level of the conventional photo-multiplier.

INTRODUCTION

In the field of soft x-ray spectroscopy it would be advantageous if improved techniques could be developed for the purpose of measuring relative intensities. The spectral region of interest in this connection lies below 500A. To date, the detection of radiation in this wave-length region has been confined to the use of special photographic emulsions such as Schumann and Ilford Q plates. In conventional vacuum spectroscopy such plates are satisfactory for wave-length determinations. In the study of soft x-ray emission or absorption bands the investigation is usually concerned with the intensity distribution of the radiation. To the existing troublesome features of vacuum spectroscopy, the methods of photographic photometry introduce additional laborious procedures which at best are subject to relatively large errors. The primary purpose of this research was to investigate the feasibility of using photo-multiplier tubes for the detection of radiation in the spectral range of 50 to 500A with the hope that this device might be useful for the rapid comparison of intensities in the extreme ultraviolet. The photo-multiplier has been used successfully for similar purposes in the visible region.[1]

Since the extreme ultraviolet radiation is completely absorbed by the glass envelope of the multiplier tube, it is necessary to convert the incident radiation by means of a phosphor. The light produced in the phosphor sample may then be utilized to illuminate the photo-cell. This method has been applied to the measurement of light output by luminescent screens irradiated by hard x-rays.[2] A survey of the literature for the purpose of selecting a suitable phosphor yielded meager information[3] on the response of inorganic phosphors in the soft x-ray region. Consequently, an investigation was undertaken to determine whether or not the common phosphors could be excited by primary radiation in the wave-length range 50 to 500A.

In carrying out measurements of relative intensity, it is of importance to know whether the phosphor sample used in conjunction with the multiplier photo-cell can be relied upon to give a faithful reproduction of intensity variations present in the incident radiation. After the selection of a suitable luminescent material, this aspect of the conversion process was tested by comparing the distribution in intensity of the light emitted by a particular phosphor sample with that of the exciting radiation. The discussion which follows is a progress report on the preliminary phases of the research and is by necessity semi-quantitative in nature.

EXPERIMENTAL

For the major portion of the work a grazing incidence vacuum spectrograph was used, though some preliminary tests were carried out in the neighborhood of 1000A using a normal incidence spectrograph. The dispersing element was a lightly ruled concave glass grating with 30,000 lines per inch and a radius of curvature of 1.5 m. The angle of incidence was 84.3°, resulting in observable wave-lengths as low as 60A. The dispersion was 1.0A/mm at 100A. The actual resolving power as computed directly from observation made with the instrument was 800 at 200A in the first order.

For the tests contemplated, the primary source of radiation ideally should possess high intensity with lines or bands distributed over the spectral region of interest. Furthermore, it should be possible to vary the source intensity in a known controllable manner. For the wave-length range under consideration, it is difficult to find a source that meets these requirements fully. The hollow-cathode helium tube

[1] G. H. Dieke, J. Opt. Soc. Am. **35**, 471 (1945).
[2] G. Fonda and F. Seitz, *Solid Luminescent Materials* (John Wiley & Sons, Inc., New York, 1948), p. 442.
[3] T. Suga and M. Kamiyama, "Photography by luminescent screens in the extreme ultraviolet," J. Opt. Soc. Am. **31**, 592 (1941).

and the condensed spark discharge between two metal electrodes or in a glass capillary, yield intense radiation in the ultraviolet. In the present experiments, the latter source, a low pressure air discharge in a glass capillary, was used in the preliminary trials while attempting to excite luminescence in the different samples of phosphors. The discharge was operated at the rate of one breakdown every 2 seconds, each spark dissipating the energy in a 1.2-μf condenser charged to 10 kv. From 20 to 80 flashes were found to be adequate for most exposures.

A second type of primary radiation available in this region consists in the characteristic soft x-ray bands, which are emitted by the light metals (Be, Mg, Al) under low voltage electron bombardment. The relative intensity distribution within such bands has been determined rather accurately.[4] Unfortunately, with the available experimental arrangements such radiations are of low intensity and their production requires the maintenance of high vacua ($\sim 10^{-6}$ mm of Hg) and freedom from impurities. In connection with the study of the response of a particular phosphor as a function of incident intensity, the sample was irradiated by the characteristic L emission band of aluminum. The radiation was generated in an x-ray tube whose target was coated with aluminum by evaporation. When operated at a potential of about 1–2 kilovolts and tube current of 200 ma, exposure times ranging from 1 to 30 hours were required for the various tests.

Since the radiation diffracted from the grating is brought to a focus on the Rowland circle, the phosphor layer had to be deposited on a flexible backing strip which could then be mounted vertically on the plate holder track (see Fig. 1). In the deposition of the phosphor layer, the use of binders must be reduced to a minimum, since the incident radiation is absorbed highly by such materials. For example, coating a wet gelatin surface proved to be unsatisfactory. From the standpoint of efficient conversion and reproducibility the best results were obtained when the desired thickness of the phosphor was allowed to settle on a strip of celluloid from a water suspension. Such a coating, however, was not sufficiently durable when long exposures were involved. More permanent samples were made either by dusting on a glass slide painted with sodium silicate or by coating the slide with a film of organic binder (amyl acetate) containing the luminescent substance. In both instances the specimens were baked for several hours at 400°C. The phosphor thickness used fell in the range of 3 to 12 mg/cm^2. Under comparable irradiation the light output of the samples prepared with the silicate and organic binders were, respectively, about 1/3 and 1/10 as much as that of the samples obtained by settling from a water suspension.

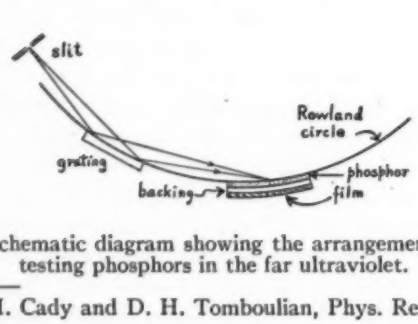


FIG. 1. Schematic diagram showing the arrangement used in testing phosphors in the far ultraviolet.

The light emitted by the phosphor was detected photographically by placing strips of sensitive film in contact with the uncoated face of the transparent backing material. Eastman ortho- or panchromatic film was used depending on the expected spectral distribution of the light emitted by the particular phosphor.

RESULTS AND DISCUSSION

The various phosphors tested and the results obtained are listed in Table I. All samples were deposited on celluloid strips from water suspension and were treated alike as to exposure (75 sparks) and development. The relative intensities of the converted radiation were determined photometrically for the exciting radiation of 172A. Since $CaWO_4$:Pb was observed to have the greatest sensitivity, the response of this phosphor was chosen as the basis for comparison. The relative response of the phosphors listed remains sensibly the same for other primary wave-lengths extending over a range of 50A above and below the primary radiation at 172A.

In the spectral range investigated, the energy of the excitant photons lies in the 50- to 100-ev range. Photons of this energy are transformed to lower energy quanta (of the order of a few ev). In general, the conversion efficiency—total light energy generated divided by the energy absorbed—diminishes as the energy difference between absorbed and emitted photon increases.[5] Unfortunately, it is not possible to calculate the conversion efficiency for in-going 100-ev photons since this involves the measurement of the absolute intensity distribution of the weak incident radiation—a very difficult task in this spectral region. For 80-kev x-ray photons conversion efficiencies of 5 percent have been reported.[6]

The tabulation brings out the observation that in this wave-length region there is a large difference

[4] W. M. Cady and D. H. Tomboulian, Phys. Rev. **59**, 381 (1941).

[5] G. Fonda and F. Seitz, *Solid Luminescent Materials* (John Wiley and Sons, Inc., New York, 1948), p. 159.

[6] Coltman, Ebbighausen, and Altar, J. App. Phys. **18**, 530 (1947).

in the efficiencies of $CaWO_4$ and ZnS, although these phosphors have comparable efficiencies in the visible, where the photon penetration is only 10 times as great and is in either case much greater than the spacing of luminescent centers in ZnS.

The Effect of Screen Thickness

In the case of the fluorescent screen used with hard x-rays, light is generated throughout the body of the thick phosphor layer (150μ) which is a discontinuous assemblage of crystals of various sizes (0.01–1.0μ). Luminescent centers are activated at various depths. The light from each center is reflected, refracted, and absorbed by individual crystals and eventually arrives at the front or the back surface of the layer. Depending upon the absorption coefficient of the material for x-rays and visible light, there is an optimum thickness of phosphor at which the light output reaches a maximum. For the photon energies dealt with in this work the situation is somewhat different in that the penetration of the incident beam does not exceed a few microns, and the luminescence originates essentially at the surface of the phosphor layer which is irradiated. As a consequence, approximately one-half of the total light is lost to the surroundings at this surface and a portion of the remainder is transmitted to the photographic film. Since visible light was weakly absorbed by the phosphors tested, observations confirmed the expectation that the thickness of the phosphor layer should not be a critical factor in the light output. Two samples of $CaWO_4$, approximately 10μ and 150μ in thickness, essentially gave rise to the same photographic density of recording film under equal excitation.

Distortion of the Image

The luminescent radiation originates essentially on the surface of the phosphor layer from a well-defined area corresponding to the image of a spectrum line. However, because of scattering by phosphor grains, non-uniform distribution of excited centers in the layer, and the fact that the photographic film is not in contact with the luminous region of the screen, the resultant image as photographed on the film is distorted. The spreading of the image resulting from all causes was investigated by comparing the shape of a monochromatic line $\{\lambda = 150\text{A } (OV)\}$ as focused on the screen with the resultant shape of the same line after conversion by the phosphor. Curve (a) in Fig. 2 shows the relative intensity distribution of the line as a function of distance along the Rowland circle. The curve was determined photometrically by the use of Ilford Q plates. The relative intensity distribution in the image of the same line after the passage of the light through the phosphor and backing layers was obtained from a calibration of the ortho film used to photograph the luminescent light. All instrumental slit widths were kept constant in these determinations. Assuming that the conversion process does not alter the relative distribution in intensity of the incident light and neglecting the effects resulting from absorption, scattering, and nonuniformity of the phosphor deposit, it is possible to calculate the spreading of the converted image caused by the thickness of the backing strip. For this calculation the primary image is considered as a diffuse luminous surface, and as a further simplification the original distribution is replaced by one which has a square top and vertical sides. If the width of the rectangular distribution is $2a$, and if h is the thickness of the backing film, the intensity I along a median line of the resultant image on the film is given by

$$I = k\left[\frac{(a-l)}{(h^2+(a-l)^2)^{\frac{1}{2}}} + \frac{(a+l)}{(h^2+(a+l)^2)^{\frac{1}{2}}}\right],$$

TABLE I. The response of various fluorescent materials.

Phosphor	Approx. wave-length of max. emission in Angstroms	Relative fluorescent intensity
Tungstates:		
tetr. $CaWO_4$:Pb	4400	1.00[a]
tetr. $CaWO_4$:[W]	<4300	0.70
monocl. Mg_2WO_5:[W]	4800	0.60
monocl. $CdWO_4$:[W]	4930	0.18
Silicates		
rbhdrl. Zn_2SiO_4:Mn	5200	0.15
$Ba_2Si_4O_{11}$:Pb	?	0.0
Others:		
$BaSO_4$:$PbSO_4$	4100	0.85[a]
$Ca_3(PO_4)_2$:Ce	3600	0.15
ZnO:[Zn]	5050	0.15[b]
$Cd_2B_2O_5$:Mn	6150	0.0
tub. ZnS:Ag	4550	0.0[b]
Hex. ZnS:CdS:Ag	5700	0.0[b]

[a] Also tried at 1000A. Good conversion.
[b] Also tried at 1000A with panchromatic film. No observable response.

TABLE II. Relative efficiencies for different source intensities.

Primary wave-length	Source	Incident intensity[a]	Converted intensity	Intensity converted/incident	Rel. efficiency spark/x-rays
172A	spark	$2.17K$	$0.586K'$	$0.27K'/K$	0.78
172A	x-rays	$3.92\times10^{-4}K$	$1.38\times10^{-4}K'$	$0.35K'/K$	
192A	spark	$2.24K$	$0.715K'$	$0.32K'/K$	0.92
192A	x-rays	$3.56\times10^{-4}K$	$1.28\times10^{-4}K'$	$0.35K'/K$	
				Average	0.85

[a] The intensities are in arbitrary units. K and K' represent unknown calibration factors for the two emulsions by means of which the observed densities could be converted to photons/cm²-sec.

where k is the brightness of the screen, and l is a distance measured from the peak of the resultant image. In this experiment $2a=0.58$ mm and $h=0.585$ mm. The assumed incident distribution, the geometrically predicted shape, and the observed line shape are shown respectively in curves (b), (c), and (d) of Fig. 2. The observed and calculated curves are plotted by matching the maximum ordinates arbitrarily. The spreading of the image is thus seen to be largely caused by the intervening thickness of the backing layer and is primarily a result of the mounting scheme used in these tests. If a narrow movable slit be utilized to scan the image on the phosphor layer, and if the light from the slit be detected by means of a photo-tube, patterns corresponding to the initial line shape may readily be obtained provided the brightness of the image is sufficiently high.

Relative Efficiency *vs.* Incident Intensity

A second part of this investigation consisted in an attempt to determine whether the efficiency of a $CaWO_4$ screen was a function of the intensity of the incident radiation. The spark discharge and the soft x-ray tube were available as two sources having intensity ratios of 5000:1. At corresponding wave-lengths in the common spectrum range of the two sources (170–200A), it was possible to compare the relative efficiency of the phosphor at the two intensity levels with Ilford Q-1 plates serving as the comparison standard for the incident radiation. A comparison of the relative intensities of the converted radiation was arrived at photometrically through the use of ortho film. The measurements at two different wave-lengths are shown in Table II. Considering the known ratio of exposures and neglecting the reciprocity failure, the relative efficiency of the phosphor for the two sources with an approximate intensity ratio 5000:1 was found to differ from unity by about 15 percent—an amount which is of the order of the experimental error involved in determining the relative sensitivity of the different types of screens used in connection with the spark and x-ray source. Hence, no significant change in efficiency of conversion is to be expected over the above intensity range.

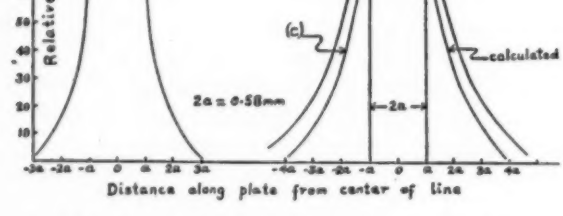


FIG. 2. Effect of luminescent screen on the shape of a spectrum line ($\lambda=150$A, OV).

As a further test of this nature, the relative intensity distribution of the L emission band of aluminum after conversion by a $CaWO_4$ screen was compared with the distribution in intensity of the incident radiation as given in curve (a) Fig. 3. The ordinate represents the relative intensity as a function of the energy of the photon which is emitted when a vacancy in the L *II–III* levels in aluminum is filled by electron transitions from the conduction band. The data for this curve were obtained from previous determinations[4] made in this laboratory. The head of the band is located at 170A, and the band extends over a region of 35A towards longer wave-lengths. After transformation, the intensity of the light emitted by the phosphor when plotted as a function of the incident photon energy is shown in curve (b) Fig. 3. The data for this curve were obtained by photometric reduction from the average of five microphotometer traces of the band photographed on Eastman ortho film. The emulsion was calibrated for blue light ($\lambda=4300$A) corresponding to the peak in the emission spectrum of $CaWO_4$. The spread in the values of the intensity obtained from each of the individual traces is of the order of 6 percent. A second and stronger exposure obtained by irradiating a different portion of the screen showed greater unsystematic variations presumably a result of non-uniformity of phosphor coating. The area utilized over the surface of the sample is a strip 6 mm wide, 25 mm long.

The incident radiation varies slowly with the photon energy over the portion of the band extending from 65 ev to 70 ev. At a given photon energy over this range the ratio of ordinates of curves (a) and (b) is a constant equal to 0.684. Multiplying the ordinates of curve (b) by this factor one obtains curve (c). This procedure is equivalent to matching the two distributions over the middle range of photon energies. Excluding for the moment the region of the peak, the relative intensity distribution of the light given off by the phosphor is essentially coincident with that of the exciting radiation.

Superimposed upon the characteristic primary radiation from the x-ray tube, there is a weak continuum whose contribution to the total intensity has been subtracted in arriving at the results shown in curve (a). A similar correction for background radiation has been applied in obtaining the result for the luminescent yield. In the latter case the background is also due to stray visible light whose

effect becomes more pronounced in the case of runs with the phosphor sample because of longer exposures and the sensitivity of the emulsion to visible light. The correction for fog introduces particularly large uncertainties in the region of low photon energies.

The lowering of the peak intensity in the converted spectrum is ascribed to the spreading of the image produced by the thickness of the supporting strip. Here the incident distribution resembles that of a discrete spectrum line and is of comparable width. The situation is similar to that discussed previously in connection with the observed alteration produced in the shape of the geometrical image of such a line. Near the head of the primary band the intensity falls to zero abruptly, and the photometric determinations in this region are not reliable. Nevertheless, a widening of the primary band as seen in curve (c) of Fig. 3 was observed in all individual reductions.

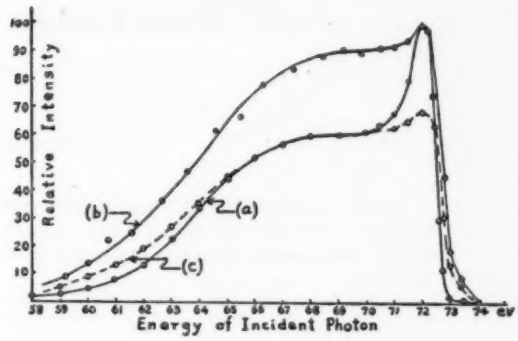


FIG. 3. The intensity distribution in the L emission band of aluminum upon conversion by $CaWO_4$ screen.

The Use of Photo-Multiplier

The normal incidence spectrograph was modified for the purpose of testing the photo-multiplier as a detector of ultraviolet radiation in the 1000A region. The conventional plate holder was replaced by an opaque brass plate containing a slit, and a phosphor coated strip of celluloid was mounted back of this occulter in the focal plane of the spectrograph. The housing for the photo-multiplier tube (931A) was so designed that only the light emitted by the phosphor could reach the cell. The voltage dividing resistors were located directly under the tube base and the entire assembly was placed within the vacuum chamber. High voltage leads were admitted through insulated electrodes sealed in the removable end cover of the spectrograph chamber. Provision was made for rotating the grating through a small angle about a vertical axis in order to illuminate the plate holder slit by different spectrum lines. Using the glass capillary discharge as a source, preliminary trials indicated that the converted radiation was detected by the tube. Further refinements in instrumentation or measurements were not undertaken for this spectral region, since the rough tests established the possibility of this type of detection with strong sources and efficient phosphors.

Corresponding tests with much weaker x-ray sources in the extreme ultraviolet have not been carried out as yet. A fifteen-hour exposure of the orthoemulsion using a $CaWO_4$ screen yields a density which corresponds to an output intensity of 1.5×10^5 photons of blue light per cm^2 per second. Taking the limit of detectability by the photo-multiplier at room temperature to be about 2×10^4 photons per cm^2 per sec.[7] it appears that the available intensity is above this limit by only a factor of ten. This is an uncomfortably narrow margin. Redesign of spectrograph, x-ray tube, and cooling of the photo-multiplier tube would improve matters by a factor of 50. It is, therefore, justifiable to state that with complicated though feasible techniques the photo-multiplier tube using a $CaWO_4$ converter would serve as a detector for soft x-ray radiation in the far ultraviolet.

The authors are greatly indebted to Dr. H. W. Leverenz of RCA who supplied the phosphors and wish to acknowledge gratefully the assistance of Dr. J. H. Webb of Eastman Kodak Company who furnished the calibration data for the orthoemulsion used in certain tests. The research was supported by the Office of Naval Research under contract N6ori-91.

[7] R. W. Engstrom, J. Opt. Soc. Am. 37, 420 (1947).

Phase Lead for A.C. Servo Mechanisms

ARTHUR H. BENNER
Ordnance Research Laboratory, State College, Pennsylvania

(Received August 31, 1948)

General design curves for use of lead networks on log-db plots are given. A variety of parallel and bridge T networks along with d.c. networks are evaluated for use in a.c. servo mechanisms. The restrictions existing on some networks are enumerated, and suggestions on general design procedure are made.

IT is the purpose of this paper to evaluate the different methods of obtaining lead control in a.c. carrier servo mechanisms.

GENERAL TECHNIQUE

A convenient method of analysis of the performance of a servo system is the "log-db" method. In this, the transfer function of the servo, $\theta_0/\epsilon = KG$, is plotted in two parts; $|\theta_0/\epsilon|$ db *vs.* ω and ϕ *vs.* ω. The magnitude and phase of the transfer function of a lead network cascaded with the servo system may be added to the plots of the uncorrected servo to obtain the corrected response. From these parametric curves and a phase margin plot,[1] the important characteristics of the servo mechanism such as the natural frequency ω_n, the cut-off frequency and phase margin at gain crossover, the gain constant, and the gain margin at phase crossover may be determined. Acceptable values of these factors depend largely on the particular servo system involved.[1-3]

FIG. 1. ϕ *vs.* ω/ω_1 for $\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j(\omega/\omega_1)\alpha}{1+j(\omega/\omega_1)}$.

Most types of lead networks used for both d.c. or a.c. servo mechanisms may be expressed by the transfer function[4]

$$\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+T\alpha p}{1+Tp}, \tag{1}$$

where α is the attenuation and T the network time constant. In real frequencies this is often expressed as

$$\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j(\omega/\omega_1)\alpha}{1+j(\omega/\omega_1)}, \tag{2}$$

where $\omega_1=1/T$. For analysis on log db plots, Eq. (2) may be broken into its magnitude and phase as a function of ω/ω_1 as

$$\left|\frac{e_0}{e_1}\right| db = 20\log\left(1+\left(\frac{\omega}{\omega_1}\alpha\right)^2\right)^{\frac{1}{2}} - 20\log\left(1+\left(\frac{\omega}{\omega_1}\right)^2\right)^{\frac{1}{2}} - 20\log\alpha, \tag{3}$$

$$\phi=\tan^{-1}\frac{\omega}{\omega_1}\alpha-\tan^{-1}\frac{\omega}{\omega_1}. \tag{4}$$

These last two equations are plotted for common values of α in Figs. 1 and 2 and are valuable design curves.

The maximum phase shift of a network given by Eq. (2) occurs at

$$\omega_{\phi m}/\omega_1=1/(\alpha)^{\frac{1}{2}} \tag{5}$$

[1] James, Nichols, and Phillips, *Theory of Servomechanisms* (McGraw-Hill Book Company, Inc., New York, 1947).
[2] A. C. Hall, *The Analysis and Synthesis of Linear Servomechanisms*, Technology Press, M.I.T. (May 1943).
[3] Lauer, Lesnick, and Matson, *Servomechanisms Fundamentals* (McGraw-Hill Book Company, Inc., 1947).
[4] See reference 2, p. 96.

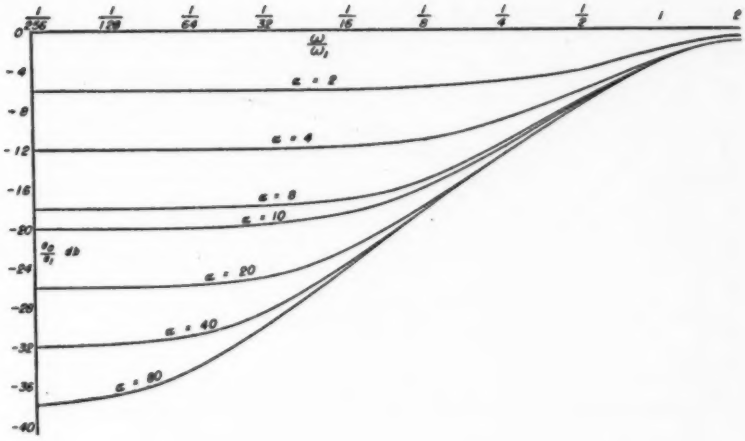


FIG. 2. e_0/e_1 db *vs.* ω/ω_1 for $\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j(\omega/\omega_1)\alpha}{1+j(\omega/\omega_1)}$.

where ω_{ϕ_m} is frequency of maximum phase shift. Equation (5) is plotted in Fig. 3a. This expression is important because it can be shown that in general, the maximum phase shift of a lead network should be placed at a certain frequency, depending upon the servo mechanism involved, for optimum results.

The correct choice of this frequency is dependent on the type of servo mechanism involved. Because of this fact, no definite statement can be made covering all possible cases. In general, several "rules of the thumb" may be used as a guide for the choice of this frequency of maximum phase shift, although final results usually call for a "cut and dry" graphical solution until favorable characteristics are realized. First, in a servo system whose transfer function is of the ideal second-order type, the frequency of maximum phase shift may be placed as high in the frequency spectrum as desired with corresponding improvement in the frequency response of the system, limited only by the expense of gain. Unfortunately, this ideal type of system is seldom encountered in practice. Secondly, in the more common servo mechanisms whose transfer functions are third or higher order, the frequency of maximum phase shift is often limited to that portion of the frequency spectrum in the vicinity of the natural frequency of the uncorrected servo system and several octaves about this frequency. This restriction becomes more severe, that is, the frequency spectrum becomes narrower, as the relative magnitude of these higher order terms increases. This restriction arises primarily from the necessity of holding other factors of the system to acceptable values, for example, the gain margin at phase crossover, and to minimize the distortion of the shape of the servo transfer function. Experience has shown that, in general, second-order servos and higher order systems whose higher order terms are small may have their natural frequency increase "several" octaves with lead control, as contrasted with higher order systems whose higher order terms are relatively large may have their natural frequency increased only a "few" octaves. The value of this maximum phase shift, ϕ_{max}, is

$$\phi_{max}=\sin^{-1}(\alpha-1/\alpha+1). \tag{6}$$

This equation is plotted in Fig. 3b. Examination of Fig. 3b and practical considerations dictate that little is gained for an α of a single network greater than about 20, as ϕ_{max} increases slowly with large values of α. For example, a single network with an attenuation of 40 db has a ϕ_{max} of 78° while two cascaded networks of 20 db attenuation each (properly isolated) would give a ϕ_{max} of 110°.

SPECIFIC NETWORKS

Phase lead for error rate control of an a.c. carrier servo mechanism using passive networks may be obtained by; (1) notch type a.c. network and (2) by demodulating the suppressed carrier signal and using d.c. lead networks.

As case 1, consider the network shown in Fig. 4a. The transfer function of this network is

$$\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j[(\omega_c+\omega)/2\omega][(\omega_c-\omega)/\omega_1]\alpha}{1+j[(\omega_c+\omega)/2\omega][(\omega_c-\omega)/\omega_1]},$$

where $\omega_1=\alpha/2R_1C$, $\alpha=(R_1+R_2)/R_2$, $\omega_c=1/(LC)^{\frac{1}{2}}$ the carrier frequency, $\omega_s=\omega_c-\omega$, and ω_s is the signal (modulation) frequency. If $\omega_c\gg\omega_s$ then $(\omega_c+\omega)/2\omega\approx 1$,

$$\frac{e_0}{e_1}\simeq\frac{1}{\alpha}\frac{1+j(\omega_s/\omega_1)\alpha}{1+j(\omega_s/\omega_1)}. \tag{7}$$

The only restrictions on this network are the choice of practical components and the error involved in not considering the resistance of the inductor.

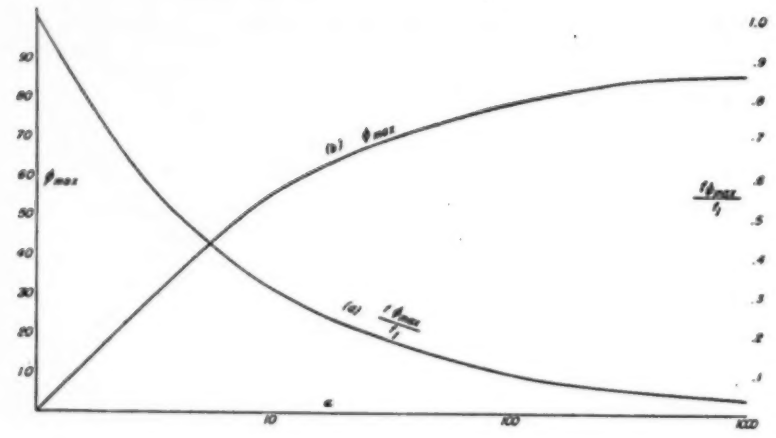


FIG. 3. (a) $\frac{f\phi_{max}}{f_1}$ vs. α (b) ϕ_{max} vs. α for $\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j(\omega/\omega_1)\alpha}{1+j(\omega/\omega_1)}$.

Equation (7) is of the same form as Eq. (2) and therefore, Figs. 1 to 3 may be used for design. If the resistance of the inductance is considered as in Fig. 4b, it follows that

$$\frac{e_0}{e_1}=\frac{R_2R_L+R_2R_1-\omega^2R_2R_1CL+j\omega R_2(L+R_1R_LC)}{R_2R_L+R_2R_1+R_1R_L-\omega^2R_2R_1CL+j\omega(LR_2+R_1R_2R_LC+R_1L)}, \tag{8}$$

and

$$\omega_e^2=\frac{R_1R_LL-R_2R_LL+R_1^2L-R_1^2R_L^2C}{L^2CR_1^2}, \tag{9}$$

which reduces to $\omega_e^2=1/LC$, if $R_L=0$. Equation (8) approximates Eq. (7) if the terms involving R_L are sufficiently small to be neglected. Let us consider a practical example and from it observe the error involved in neglecting R_L. Assume that the values of the circuit in Fig. 4a are $f_e=400$ c/sec., $\alpha=10$, $L=0.1hy$, $R_1=450K$, $R_2=50K$, $C=1.57$ μf. If, however, R_L is a reasonable value such as encountered in a toroid, say 20 ohms, from Eq. (9), f_e is then 401 c/sec. This value is small enough to be ignored. Of course, a notching filter of this type requires that the carrier frequency be held to close tolerances.

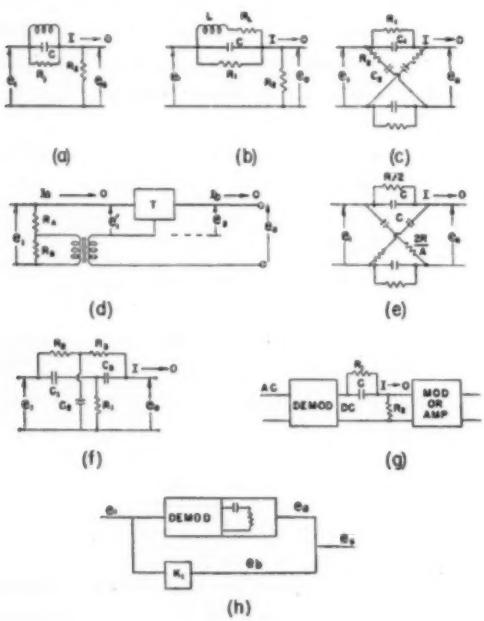


FIG. 4.

For case 2, consider the circuit in Fig. 4c. This is the lattice equivalent of a parallel T network. Its transfer function is

$$\frac{e_0}{e_1}=\frac{R_1C_1+R_2C_2-R_1C_2}{R_1C_1+R_2C_2+R_1C_2}\times\frac{1+j[\omega^2R_1R_2C_1C_2-1/R_1C_1+R_2C_2-R_1C_2]}{1+j[\omega^2R_1R_2C_1C_2-1/R_1C_1+R_2C_2+R_2C_2]}.$$

For this equation to have the characteristics of a lead network, $\omega_e^2=1/R_1R_2C_1C_2$ and $R_1C_1+R_2C_2-R_1C_2>0$. If $R_1=AR_2$ and $C_2=BC_1$, then

$$\frac{e_0}{e_1}\simeq\frac{1+(B/A)-B}{1+(B/A)+B}\times\frac{1+j(\omega_s/\omega_e)[2(B/A)^{\frac{1}{2}}/1+(B/A)-B]}{1+j(\omega_s/\omega_e)[2(B/A)^{\frac{1}{2}}/1+(B/A)+B]}. \tag{10}$$

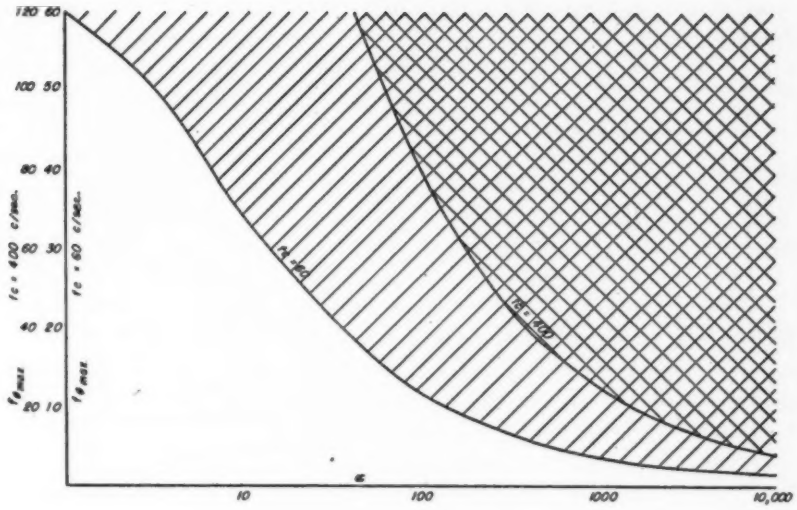


FIG. 5. $f_{\phi max} \geq \frac{2f_c(\alpha)^{\frac{1}{2}}}{\alpha+1}$.

For Eq. (10) to be similar to Eq. (2),

$$\alpha = \frac{1+(B/A)+B}{1+(B/A)-B},$$

$$\omega_1 = \frac{\omega_c}{2}\,\frac{1+(B/A)+B}{(B/A)^{\frac{1}{2}}}.$$

Solving for A and B and applying the condition that A and B must be positive for the network to be realizable, the following restrictions are found,

$$\alpha > 0,$$

and

$$f_1 > \frac{2\alpha f_c}{\alpha+1}. \tag{11}$$

Figure 5 shows a plot of $f_{\phi max} > 2f_c(\alpha)^{\frac{1}{2}}/(\alpha+1)$ in its lower limit for $f_c = 400$ and 60 c/sec. $f_{\phi max}$ must fall in the appropriate cross hatched region for a given α. From Fig. 5, it is evident that severe restrictions exist on the usefulness of this network. Fig. 5 shows that large values of α are necessary to pull $f_{\phi max}$ to low values. This indicates that these networks are rather useless for correction of servo mechanisms with a low natural frequency. This is especially true of the 400-cycle case.

Many published tables[4, 5] for design of parallel or bridge T networks make use of notch width rather than network time constant. It can easily be shown that if $\omega_s \ll \omega_c$ and α is large, the notch width, n, in radius per second is

$$n = 2/T\alpha,$$

where n is defined as the frequency difference between the points on the filter characteristic whose magnitude is $\sqrt{2}$ times the minimum output. Hence

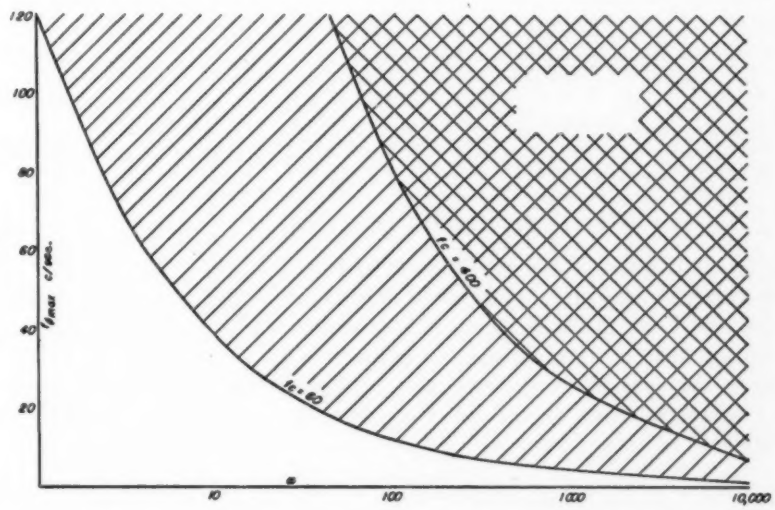


FIG. 6. $f_{\phi max} \geq 2f_c/(\alpha)^{\frac{1}{2}}$.

[5] A. Sobczyk, *Parallel T Stabilizing Networks for* A-C *Servos*, RL Report No. 811 (March 7, 1946).

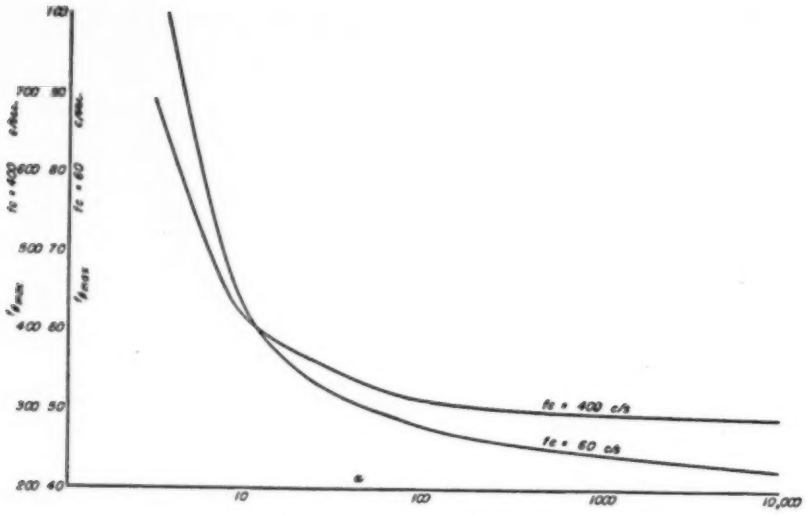


FIG. 7. $f_{\phi max}=\frac{(\alpha)^{\frac{1}{2}}f_c}{\sqrt{2}(\alpha-1)^{\frac{1}{2}}}$.

for this network

$$n>2\omega_c/(\alpha+1)$$

for physical realizability.

In case 3, a lattice is again considered, but whose theoretical response at ω_s is zero. Its configuration is the same as that in Fig. 4c, but the conditions are different as $R_1C_1+R_2C_2-R_1C_2=0$. Then if $R_1=\zeta R_2$,

$$\frac{e_0}{e_1}=\frac{[(\zeta-1)^{\frac{1}{2}}/\zeta]j(\omega_s/\omega_c)}{1+j(\omega_s/\omega_c)[(\zeta-1)^{\frac{1}{2}}/\zeta]}. \tag{12}$$

This circuit may conveniently be altered into a parallel T network and used in a circuit as shown in Fig. 4d. If I_a and I_b are taken as zero,

$$\frac{e_0}{e_1}=(1-K_1)\frac{1+j(\omega_s/\omega_c)[(\zeta-1)^{\frac{1}{2}}/(1-K_1)\zeta]}{1+j(\omega_s/\omega_c)[(\zeta-1)^{\frac{1}{2}}/\zeta]}, \tag{13}$$

where $R_a/R_a+R_b=K_1$.

This is similar to Eq. (2) if

$$\alpha=1-K_1,$$

and

$$\omega_1=\omega_c\zeta/(\zeta-1)^{\frac{1}{2}}. \tag{14}$$

Since $\zeta>0$, $0<\omega_c/\omega_1<\frac{1}{2}$, then

$$f_{\phi max}=(f_1/(\alpha)^{\frac{1}{2}})\geq(2f_c/(\alpha)^{\frac{1}{2}}). \tag{15}$$

Equation (15) is plotted in the lower limit in Fig. 6 for $f_c=400$ and 60 c/sec. These curves also indicate that this network is of little practical value for servo systems whose natural frequency is low.

For case 4, the lattice equivalent of a bridge T circuit is considered and shown in Fig. 4e. The transfer function of this circuit is

$$\frac{e_0}{e_1}\simeq\frac{2}{A+2}\frac{1+j(\omega_s/\omega_c)(A)^{\frac{1}{2}}}{1+j(\omega_s/\omega_c)(2(A)^{\frac{1}{2}}/A+2)}. \tag{16}$$

This equation is in the form of the general Eq. (2), where

$$\alpha=\frac{A+2}{2},\quad \omega_1=\frac{(A+2)\omega_c}{2(A)^{\frac{1}{2}}},\quad \omega_c^2=\frac{A}{R^2C^2}.$$

In this network α and ω_1 are not independent, that is, for a given α, ω_1 is fixed. The frequency of maximum phase shift of this network is

$$f_{\phi max}=\frac{(\alpha)^{\frac{1}{2}}f_c}{\sqrt{2}(\alpha-1)^{\frac{1}{2}}}. \tag{17}$$

Equation (17) is plotted in Fig. 7 and shows that again these networks are of little value for correction of systems whose system natural frequency is low.

For case 5, the asymmetrical parallel T described by Sobczyk[3,5] is treated. Following his notation and circuit as shown in Fig. 4f, the transfer function is

$$\frac{e_0}{e_1}=\frac{L}{T_d}\frac{1+jT_d(\omega+\omega_c/2\omega)(\omega-\omega_c)}{1+jL(\omega+\omega_c/2\omega)(\omega-\omega_c)}. \tag{18}$$

Where ω_s is small, Eq. (19) may be approximated by

$$\frac{e_0}{e_1}\simeq\frac{L}{T_d}\frac{1+jT_d\omega_s}{1+jL\omega_s}. \tag{19}$$

This equation is of the form of Eq. (2) where

$$1/\alpha=L/T_d,\quad L=1/\omega_1.$$

Sobczyk has shown that for the network to be realizable $L\leq 1/\omega_c$, where ω_c is the carrier frequency. Hence $\omega_1>\omega_c$ and

$$f_{\phi max}\geq f_c/(\alpha)^{\frac{1}{2}} \tag{20}$$

which is plotted in Fig. 8 for $f_c=400$ and 60 c/sec.

This network, as did cases 2 through 4, exhibits very little phase shift at low frequencies without the expense of great attenuation. If this attenuation can be afforded, the network can be used to correct even a low frequency system.

The last case, 6, is that involving demodulation and use of d.c. lead networks. This case will be divided into two parts, (a) single channel and (b) dual channel.

Case 6(a) is illustrated by Fig. 4g. It consists of demodulating the a.c. carrier, operating on the demodulated wave with a derivative plus proportional network, and a modulator to restore the signal to the a.c. carrier level or a d.c. amplifier if it is not desired to have the signal as suppressed carrier a.c. The network in this system has the familiar general transfer function

$$\frac{e_0}{e_1}=\frac{1}{\alpha}\frac{1+j(\omega_a/\omega_1)}{1+j(\omega/\omega_1)},$$

where $\alpha=(R_1+R_2)/R_1$ and $\omega_1=1/R_1C$. There are no restrictions on this system and design curves 1 through 3 may be used.

Case 6(b) is illustrated by Fig. 4h. Such dual channel systems are often desirable when other signals are present in the input that are not to be operated on by a lead network (repeat back selsyns, tachometers, etc.). The transfer function of this system is

$$\frac{e_0}{e_1}=k_1\frac{1+j(\omega/\omega_1)(1+k_1/k_1)}{1+j(\omega/\omega_1)}, \qquad (21)$$

where $\omega_1=1/RC$. To use the design curves 1–3, Eq. (23) must be altered to compare with Eq. (2),

$$\frac{e_0'}{e_1'}=\frac{e_0}{e_1}\frac{1}{1+k_1}=\frac{k_1}{1+k_1}\frac{1+j(\omega/\omega_1)(1+k_1/k_1)}{1+j(\omega/\omega_1)},$$

where $\alpha=(k_1+1)/k_1$.

This equation may employ the design curves 1 to 3, but the value of $|e_0'/e_1'|$ db must be diminished by 20 $\log(1+k_1)$ to give $|e_0/e_1|$ db.

CONCLUSIONS

From the preceding examples, it is evident that parallel or bridge T filters are not as flexible as are $L-C$ a.c. networks or d.c. networks.

(1) It can be said that the usability of networks depend primarily on the servo to which they are applied. A second-order servo mechanism can be improved in frequency response considerably by any of the above treated networks. Similarly, it is evident that a third and higher order system whose higher order terms are small may also be improved by these networks.

(2) In third or higher order systems whose higher order terms are important, realizable parallel or bridge T networks will not present sufficient phase lead to be usable at low frequencies where it is necessary without great attenuation.

(3) It is also evident that the higher the carrier frequency of the servo mechanism, the less useful are the bridge and parallel T networks.

(4) Finally, restrictions on carrier frequency and component tolerances are much more severe on the a.c. than on d.c. networks.

(5) A designer, after examination of the θ_0/ϵ function of a given servo, may refer to curves 5 through 8 to evaluate the feasibility of parallel or bridge T networks for phase lead. If he finds a particular network suitable, subject to the restrictions evident in the curves, he may then use curves 1 through 3 for design in the log-db method.[1] If, however, the parallel or bridge T networks are not usable, he may use the more general d.c. networks or the network shown in Fig. 4a.

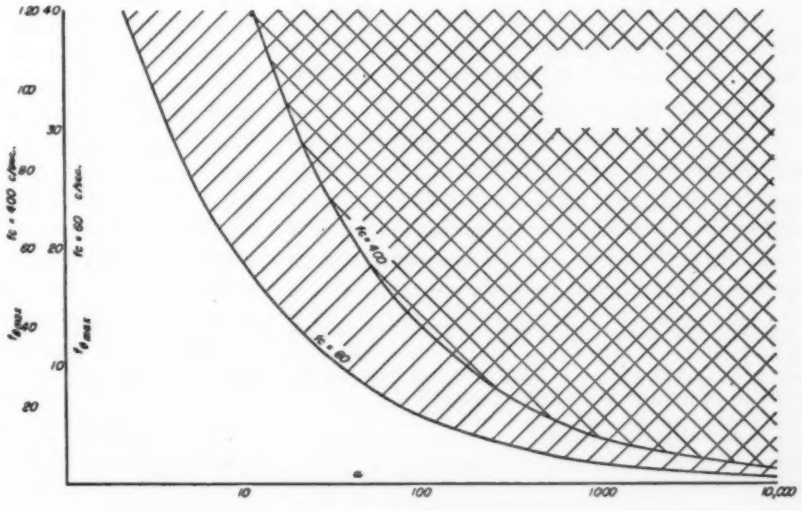


FIG. 8. $f_{\phi max} \geq f_c/(\alpha)^{\frac{1}{2}}$.

A Grain Boundary Model and the Mechanism of Viscous Intercrystalline Slip*

T'ING-SUI KÊ
Institute for the Study of Metals, The University of Chicago, Chicago, Illinois

(Received September 13, 1948)

A study of the activation energy associated with the viscous intercrystalline slip shows that the conventional theories of grain boundary, e.g., the intercrystalline amorphous cement theory and the abrupt transitional theory, are both untenable. A grain boundary model is described in which the transition region at the boundary is considered as consisting of numerous disordered groups of atoms or diffused holes. The intercrystalline slip occurs through the atomic rearrangement by thermal agitation within each "disordered group" by a shear process involving as units of flow only a few atoms. This grain boundary model and slip mechanism are consistent with experimental facts and furnish, furthermore, a unified viewpoint as to the mechanism of the viscous intercrystalline slip, the volume diffusion in metals, and the constant rate creep of metal crystals under small stress. Further experiments are described concerning the influence of previous deformation and impurities on grain boundary viscosity. It has been found that the grain boundary viscosity is lower in a specimen subjected to a heavier deformation prior to its recrystallization. A very small amount of impurities was found to be able to block partially or completely the grain boundary slip in aluminum, iron, and copper. These observations are readily understood on the basis of the proposed grain boundary model and slip mechanism.

I. INTRODUCTION

IN a polycrystalline metal the atoms of neighboring grains cannot, in general, fit together in crystallographic order, and there must be some modification of the lattices where they join. There are two principal theories as to the nature of this modification. One may be called the intercrystalline amorphous cement theory which implies, in its original form, that a separate second phase is sandwiched between two grains. This sandwiched layer is structureless and has a thickness of at least some tens of atoms disposed at random.[1] The other theory may be called the abrupt transitional theory, and assumes that the boundary is merely the surface of contact of two crystals fitting together as well as their different orientations allow. The abrupt transitional theory implies that the transition at the boundary is sudden, since the single crystal was considered as a limiting case of polycrystalline metal when the angle through which the transition is made decreases to zero.[2] The question concerning the nature of the modification cannot be easily settled by direct experiments since there is no method known at present of determining the grain boundary structure.

Through quantitative observations upon the temperature variation of internal friction and other related phenomena, we have demonstrated that grain boundaries do behave in a viscous manner.[3–6] A study of the activation energy associated with the viscous slip along grain boundaries contributes to our understanding toward grain boundary structure.[6,7] The activation energy associated with the viscous intercrystalline slip has been found to agree with that of volume diffusion in the case of alpha-brass, alpha-iron, and aluminum. Since the activation energy depends on the local order of the atoms through which the process concerned takes place, this observation indicates that, at least for these metals and probably for all metals, the local structure of the boundary region responsible for the viscous intercrystalline slip is similar to the local structure of the interior of the grains responsible for volume diffusion. Although the exact mechanism for volume diffusion in metals is not known, we can assume that, in a majority of cases, the volume diffusion occurs through some kind of imperfections in the crystal lattice.[8] The existence of a definite value of activation energy for volume diffusion indicates that the structures of these "imperfections," distributed throughout the crystal lattice, are similar to each other. We can, therefore, consider each imperfection as a separate unit similar to the concept of vacant lattice sites or holes commonly used in the study of ionic crystals. If these imperfections have a local structure and can be considered as separate units, then the regions between such "imperfection units" should be somewhat perfect and have a regular lattice structure. If the local structure of the boundary region responsible for the viscous intercrystalline slip is similar to the kind of local imperfections described above, then we can only assume that the grain boundary structure is somewhat inhomogeneous and consists of regions of order and disorder. This is irreconcilable with the amorphous cement theory which

* This research has been supported by ONR (Contract No. N-6ori-20-IV).

[1] W. Rosenhain and D. Ewen, J. Inst. Metals **8**, 149 (1912).
[2] B. Chalmers, Proc. Roy. Soc. **A162**, 120 (1937).
[3] T. S. Kê, Phys. Rev. **71**, 533 (1947).
[4] T. S. Kê, Phys. Rev. **72**, 41 (1947).
[5] T. S. Kê, J. App. Phys. **19**, 295 (1948).
[6] T. S. Kê, Metals Tech. (June, 1948) (Tech. Pub. 2370).
[7] T. S. Kê, Phys. Rev. **73**, 267 (1948).
[8] R. P. Johnson, Phys. Rev. **56**, 814 (1939); G. Wyllie, Proc. Phys. Soc. **59**, 694 (1947); F. Seitz, Phys. Rev., **74**, 1513 (1948).

assumes a structureless or completely disorganized amorphous layer at the boundary. It is also in contradiction with the abrupt transitional theory since, according to the picture presented above, the transition at the boundary region must be gradual.

The purpose of this paper is to describe a grain boundary model and the mechanism of intercrystalline slip based on this model. It will be shown that this proposed model and slip mechanism can correlate all the known experimental facts concerning the mechanical behavior of grain boundaries in metals.

II. A GRAIN BOUNDARY MODEL

Several interesting grain boundary models have recently been proposed, among which are the dislocation models by Burgers[9] and by Bragg,[10] and the "islands" model by Mott.[11] However, these models seem to be unable to correlate the experimental facts described in the last section. One suggestive common feature in these models is that the boundary region was assumed to be heterogeneous. Following this line of thought and taking into consideration the analysis made in the last section, a clear and practicable grain boundary model can be proposed in which the boundary region is considered as consisting of numerous lattice imperfections or atomic disorder. Each imperfection consists, in turn, of a group of atoms. Between these imperfections we have regions of good lattice fit. As we have mentioned above, the concept of "imperfections" used here is very similar to that of "holes" or "vacant lattice sites." However, the latter terms have been used in the study of ionic crystals to signify a vacant space in the lattice left by one missing ion or atom. Although this concept has been used also in the diffusion study of metals,[8] "it is perhaps possible that in a metal a vacant lattice point would not be stable but would break down into a more or less extended disordered region of mean density less than that of the crystalline metal."[12] The term "disordered atom group" or "diffused hole" will thus be used to avoid possible confusion.

The concept that grain boundary consists of numerous disordered groups of atoms can be simulated by the soap-bubble experiments, as recently shown by Bragg and Nye.[13] In their experiments, the crystal structure of a metal is represented by an assemblage of bubbles floating on the surface of a soap solution. The assemblages show clearly the disturbed region at the boundary where the bubbles have an irregular distribution. The width of this region has been found to be, in general, greater the smaller the bubbles. That the disturbed region is composed of regions of order and disorder becomes more evident as the bubble size becomes smaller.

III. MECHANISM OF VISCOUS SLIP ALONG GRAIN BOUNDARIES

Figure 1 shows schematically a somewhat disordered arrangement of atoms (after Orowan).[14] For the sake of future reference, we might call such an arrangement a "disordered group" of atoms. The stress distribution around the two shaded atoms shown in Fig. 1 is different from other places, and they can pass over one another by squeezing the atoms around them. Such a shear process is similar to that described by Eyring[15] in the case of liquids in terms of units of flow except that such a shear is possible only within a "disordered group." Let us show how such a mechanism of intercrystalline slip will give a Newtonian viscosity. The following treatment is along the general line given by Orowan in his treatment of the mechanism of viscous flow in solids.[14]

The passing of the two shaded atoms over one another requires a considerable amount of energy affected by the lattice energy of the material. This energy, which may be called the activation energy, A_i, can be supplied by thermal agitation if the temperature is high enough. In the course of thermal agitation, the surroundings of the two shaded atoms are subjected to a great number of rapidly varying states of stress, each representing a certain amount of energy. The number of these stress states representing an energy exceeding the value of A_i is proportional to $\exp(-A_i/kT)$, where k is Boltzmann's constant, and T the absolute temperature. If the local state of stress changes ν times per second, the frequency of local thermal fluctuation of energy in one fixed direction exceeding the amount A_i is proportional to[16]

$$\nu \exp(-A_i/kT). \tag{1}$$

An atomic rearrangement is accomplished once the energy exceeds the value of A_i, and this produces a local stress or stress release resulting in a small amount of slip.

According to the grain boundary model we proposed, there are numerous "disordered groups" of atoms in the grain boundary region. If the atomic

[9] J. M. Burgers, Proc. Phys. Soc. **52**, 23 (1940); Proc. Kon. Ned. Akad. v. Wetensch., Amsterdam **42**, 293 (1939); see also W. G. Burgers, *ibid.* **50**, 595 (1947).
[10] W. L. Bragg, Proc. Phys. Soc. **52**, 54 (1940).
[11] N. F. Mott, Proc. Phys. Soc. **60**, 391 (1948).
[12] N. F. Mott, Trans. Faraday Soc. **34**, 859 (1938).
[13] Lawrence Bragg and J. F. Nye, Proc. Roy. Soc. **A190**, 474 (1947).
[14] Ing. E. Orowan, Proc. West of Scotland Iron and Steel Inst. (February, 1947).
[15] H. Eyring, J. Chem. Phys. **4**, 283 (1936).
[16] M. Polanyi and E. P. Wigner, Zeits. f. physik. Chemie **139**, 439 (1928).

rearrangements involve only a few atoms at a time, then it is to be expected that the activation energies for the atomic rearrangement in various "disordered groups" will be densely grouped around a mean value A. The macroscopic shear strains resulting in the atomic rearrangement in various groups are also to be expected to have a mean value γ. In the absence of stress, the macroscopic strains produced in various groups are at random directions. Accordingly there is no net flow as a whole. If a shearing stress, however small, is applied as indicated in Fig. 1, a preferential direction of flow is given by the directional influence of the applied stress. The result is thus a steadily increasing slip along the direction of stress. If the applied shear stress τ is small, then the activation energy for atomic rearrangement can be expressed as $A-b\tau$ in the direction of stress where b is a constant independent of stress. The rate of flow or the strain per unit time along the direction of stress is, thus, proportional to

$$n\gamma\nu \exp[(-A+b\tau)/kT],$$

where n is the density of disordered groups. The activation energy in opposition to the direction of stress can be expressed as $A+b\tau$, and the rate of flow along this direction is proportional to

$$n\gamma\nu \exp[(-A-b\tau)/kT].$$

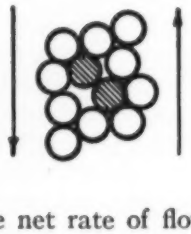


FIG. 1. Schematic illustration of a "disordered group" of atoms.

The net rate of flow along the direction of stress is thus

$$Cn \exp(-A/kT) \sinh(b\tau/kT), \qquad (2)$$

where C is a constant which has absorbed ν and γ. If the applied stress is small so that

$$b\tau \ll kT,$$

then the rate of flow is given by

$$(\text{constant}\cdot n\cdot \tau/kT) \exp(-A/kT), \qquad (3)$$

which is proportional to the stress τ, and we have a Newtonian viscosity or a "pure" viscous slip. The coefficient of viscosity is given by

$$\eta = (\text{constant}\cdot kT/n) \exp(A/kT). \qquad (4)$$

This shows that the temperature variation of η is primarily through the term $\exp(A/kT)$.

IV. GRAIN BOUNDARY SLIP, VOLUME DIFFUSION, AND CONSTANT RATE CREEP OF SINGLE CRYSTALS UNDER SMALL STRESS

Now let us see how the proposed mechanism of grain boundary slip in terms of "disordered groups" fits to experimental results as to the mechanical behavior of grain boundaries. Equation (3), which was derived on the basis of this mechanism, gives the observed viscous behavior of grain boundaries under a small shearing stress, and Eq. (4) concurs with the observed fact that the temperature variation of grain boundary viscosity follows an exponential law. According to this model the viscous slip along a plane grain boundary should be unlimited under a constant applied stress if not blocked by the interlocking effect of grain edges and corners. The extent that grain boundary slip can occur in a bi-crystal has been recently demonstrated in the case of tin by King, Cahn, and Chalmers[17] in constant load experiments. The maximum displacement they recorded so far is 0.1 mm, with an initial shear stress component of 590 g/cm² at a temperature of 222°C, which is a few degrees below the melting point of tin.

We have mentioned in Section I that in a number of metals and alloys the activation energy associated with grain boundary slip is identical, within experimental error, to that for self- or volume-diffusion. This led to the belief that the elementary units responsible for grain boundary slip are similar to those responsible for volume diffusion. From Fig. 1 we can see that the activation energy for the two shaded atoms to pass over one another depends on the local ordering around these two atoms, and this depends, in turn, on the lattice energy of the material. Although we have numerous disordered groups, the activation energies for these different groups can be considered as densely grouped around a mean value. It is this mean value of activation energy that was determined in internal friction and related experiments which are macroscopic observations. It is conceivable that the "disordered groups" of atoms we described above exist also in the interior of the grains, except that they are much smaller in number. If we assume that volume diffusion occurs through the atomic rearrangement within these "disordered groups" in the interior of grains, then the activation energy involved is necessarily the same as in the case of grain boundary slip.

Another observation is that the activation energies associated with grain boundary slip in a number of metals are also identical with those for the "constant rate creep" of metal crystals under very small stress.[7] This identity is a necessary

[17] R. King, R. W. Cahn, and B. Chalmers, Nature **161**, 682 (1948).

consequence if we assume that the mechanism of constant rate creep under small stress is similar to that of volume diffusion under the directive influence of the applied stress. A somewhat similar viewpoint has been given by Kanter.[18] The constant rate creep in single crystals can thus be considered as resulting from the atomic rearrangement of the disordered groups in the interior of grains. The essential difference between grain boundary slip and constant rate creep in crystals is that the number, n, of such disordered groups in the interior of annealed crystals is much smaller. The creep rate of single crystals should, therefore, be smaller than the creep along grain boundaries at the same temperature (Eq. (3)). Hence we need to go to a higher temperature to have a given creep rate in the case of the interior of grains than the case of grain boundaries. This has been shown to be the case in internal friction experiments in which the internal friction caused by flow in the interior of grains occurred at a higher temperature.[6]

V. EFFECT OF PREVIOUS DEFORMATION

Let us proceed now to describe some experiments concerning the influence of previous deformation and of impurities on grain boundary viscosity and compare the results with the mechanism of grain boundary slip we proposed above.

The method of study is through internal friction measurements in torsion, the procedures of which have been described in earlier papers.[3,6] The frequency of vibration in all the measurements was about one cycle per second. The specimens used were aluminum, alpha-iron, and copper.

It has been previously demonstrated[4] that the manifestations of the grain boundary viscosity can be expressed as a function of the parameter (G.S.) $\times f \times \exp(H/RT)$, where (G.S.) is the grain size of the specimen, f the frequency of vibration, H the heat of activation, and T the absolute temperature. This assumes that the other metallurgical factors are the same. Evidently the grain boundary relaxation depends also on other factors which can change the condition at the grain boundary irrespective of what model of the grain boundary we may choose.

In order to study the influence of previous cold-work upon the viscous behavior of grain boundaries, two specimens of 99.991 percent aluminum were prepared. One was subjected to a cold-work of 34 percent reduction in area (34 percent *RA*) and then recrystallized by annealing at 450°C for 2 hours. The other was subjected to a cold work of 70 percent *RA* and then recrystallized by the same heat treatment as for the first specimen. Metallographic examination showed that the average grain diameter of the 34 percent *RA* specimen is about twice as large as that of the 70 percent *RA* specimen. The internal friction curves (*versus* temperature) for these two specimens are shown in Fig. 2: Curve *I* for 70 percent *RA* specimen and curve *II* for 34 percent *RA* specimen. It is seen that the maximum internal friction in curve *II* occurs at a temperature 65°C higher than that in curve *I*. According to the previously determined relationship[4] between grain size and the optimum temperature for internal friction, the temperature shift for a change of a factor of two in grain size can only be about 15–20°C. Since the treatments on both specimens are the same except for the amount of deformation undergone prior to recrystallization, a considerable portion of this shift of internal friction curve must be connected in some way with the previous deformation. A shift of internal friction curve toward higher temperature indicates an increase of relaxation time and thus a rise in coefficient of viscosity at a given temperature. This means that the grain boundary in a specimen subjected to less previous cold work has a higher coefficient of viscosity.

One possible explanation of this observation was that the relative crystallographic orientation of adjacent grains on the two sides of the boundary varies with the amount of previous cold work. The influence of the difference of orientation of two crystals on grain boundary viscosity is a necessary consequence of the abrupt transitional theory of grain boundary. It is well known that when a polycrystalline metal is plastically deformed, the lattice orientation in individual grains is altered toward a preferred orientation in which certain lattice directions are aligned with the principal directions of flow in the metal. Such preferred orientation exists even after the recrystallization of the metal. As the process of reorientation is gradual and is usually noticeable after the metal has been subjected to a cold-work of 33 to 50 percent reduction in area, but is not completed until the metal has received reductions of 90 percent in area or more,[19] the extent of preferred orientation in the metal after recrystallization will depend on the amount of previous deformation. When there is a highly preferred orientation in the polycrystalline specimen, the lattices in adjacent grains are more similarly oriented than in the case where there is less preferred orientation. In our present case it was thought that the extent of preferred orientation in the 70 percent *RA* specimen was much greater than that in the 34 percent *RA* specimen. However, no appreciable difference was shown by x-ray examinations on the two specimens.

[18] J. J. Kanter, Trans. A.I.M.E. **131**, 385 (1938).

[19] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), p. 381.

In terms of our concept of "disordered groups" of atoms the influence of previous deformation on grain boundary viscosity means that the previous deformation can change the amount of atomic disorder at the *new* boundaries of the recrystallized grains.[20] It is conceivable that disordered atomic

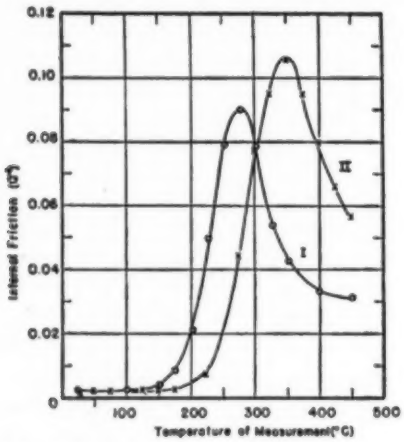


FIG. 2. Effect of previous deformation upon internal friction of polycrystalline aluminum: ⊙, 70 percent *RA* (curve *I*); ×, 34 percent *RA* (curve *II*).

groups are created in the metal during plastic deformation and that the number of disordered groups created increases with the amount of plastic deformation. When a cold-worked metal is annealed, these disordered groups migrate because of thermal agitation. Finally, a certain portion of them will reach and stay at the grain boundaries, resulting in an increase of the density of disordered groups at the boundaries. According to Eq. (3) given before, the accumulation of disordered groups at the grain boundary reduces its coefficient of viscosity. The number of disordered groups migrated to and accumulated at the grain boundary will be larger if the specimen is subjected to a larger amount of cold-work prior to recrystallization because there are more disordered groups in the grains to start with. This explains the observed effect that the grain boundaries in a specimen subjected to a larger amount previous cold work has a lower grain boundary viscosity.

VI. EFFECT OF IMPURITIES

Aluminum

The aluminum used for previous studies of grain boundary viscosity was of high purity (99.991 percent), containing as impurities 0.002 percent Si, 0.002 percent Fe, 0.002 percent Cu, and 0.003 percent Mg.** A specimen was prepared by drawing 70 percent *RA* and then annealing at 450°C for two hours. The final grain size of the specimen was about 0.03 cm. The internal friction curve (*versus* temperature) for such a specimen is shown by curve *I* in Fig. 2. A well defined internal friction peak was obtained, and the optimum temperature for internal friction maximum was about 285°C. The peak reaches a value as high as 0.09, which corresponds to a specific damping capacity of about 50 percent.

The internal friction of commercial aluminum, containing as impurities 0.18 percent Si, 0.053 percent Fe, and 0.13 percent Cu, was measured under similar conditions. Several specimens were prepared with different amounts of cold-work prior to recrystallization. They were all annealed at 450°C for two hours. As was expected, a heavier cold-working

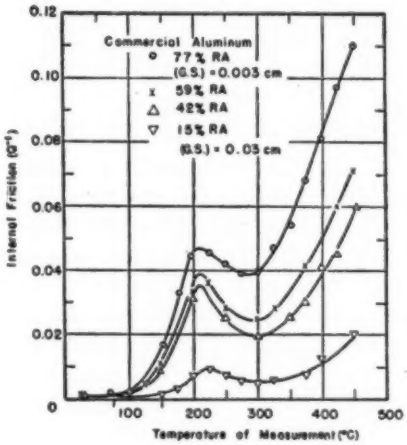


FIG. 3. Internal friction of polycrystalline aluminum of commercial purity.

gave a smaller grain size. The range of grain size obtained was from 0.003 to 0.03 cm, corresponding to the previous cold-working of from 77 to 15 percent *RA*. The internal friction curves for these specimens were determined and are shown in Fig. 3. It is rather difficult to compare each individual curve in Fig. 3 with curve *I* in Fig. 2 as they differ either in grain size or in previous deformation. However, it is certain that the internal friction peak for commercial aluminum is much smaller than that for high purity aluminum. As the total area under each peak is a measure of the amount of relaxation, we can conclude, qualitatively, at least, that the amount of grain boundary relaxation, or the extent of the viscous slip along grain boundaries, is much smaller in commercial aluminum.***

[20] An early attempt has been made to explain this effect with the concept of "lattice vacancies" or "holes," see T. S. Kê, Phys. Rev. **72**, 534A (1947).

** Information furnished by the Aluminum Company of America.

*** That impurities have the effect of decreasing the amount of relaxation of elastic modulus has been observed by Köster in the cases of aluminum and magnesium. Werner Köster, Metallkunde **39**, 9 (1948).

It is seen from Fig. 3 that the internal friction increases again at very high temperatures. This portion of internal friction is believed to have its origin in the atomic rearrangement of "disordered groups" in the interior of the grains. Such a rise of internal friction at high temperatures has also been observed in the case of magnesium,[3] alpha-iron,[7] and in copper (see Fig. 6).

Alpha-Iron

The internal friction peak associated with the grain boundary relaxation in Westinghouse Puron (a high purity iron, containing approximately 0.005 percent carbon by weight) as previously reported[6] is shown in Fig. 5 (curve (a)). The Puron was loaded with carbon, and its carbon content was raised to 0.07 percent by weight. A specimen was prepared from the carbon-loaded Puron under similar conditions as the Puron specimen, and its

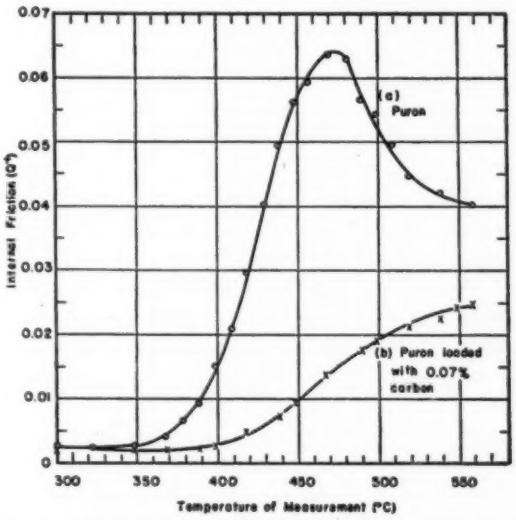


FIG. 4. Internal friction of polycrystalline iron before and after "loaded" with carbon.

internal friction curve was determined and shown by curve (b) (Fig. 4). The original high internal friction peak has now completely disappeared. Metallographic examination on these two specimens showed no perceptible change in microstructures, and there is no trace of precipitates at the grain boundaries. Evidently the amount of carbon at the grain boundaries needed to block the grain boundary slip is very small.

In this connection, it may be worth while to point out that the failure of West[21] to observe the grain boundary relaxation peak in 99.8 percent Baker's iron by elastic after-effects measurements is probably because the specimen he used contained appreciable amounts of carbon or some other impurities at the grain boundaries.

[21] W. A. West, Trans. A.I.M.E. 167, 192 (1946).

Copper

The copper specimen was prepared by drawing 99.999 percent copper to 90 percent *RA* and then annealing at 600°C in vacuum for 1½ hours. The specimen was completely recrystallized after this treatment and has a grain size of about 0.03 cm. The internal friction measurements on this specimen were taken in an atmosphere of argon purified from oxygen by filtering through hot-copper chips heated at 400°C. Curve (a) of Fig. 5 was obtained when the temperature of the specimen was increased, by steps, from room temperature up to 600°C. The specimen was then furnace-cooled. There is a definite internal friction peak around 350°C. Comparing with previous experiments on other metals, it is reasonable to believe that this peak is caused by the relaxation of shear stress across the grain boundaries in copper. When the measurements were taken again from room temperature up to 600°C, curve (b) was obtained with a much smaller peak. Similarly, curve (c) was obtained in the third trial. The specimen was then annealed at 500°C for 3 hours, and the specimen was again furnace-cooled. Curve (d) shows the internal friction curve after such a treatment, the peak of which has now almost completely disappeared. Metallographic examination of the specimen before and after the measurements shows no perceptible change in microstructures. As the specimen has not been annealed above the temperature of previous annealing, the disappearance of the peak cannot be due to the removal of some effect of cold-working during the measurement. It is also unlikely to be due to effect of grain size, as it did not change during the whole series of measurements.

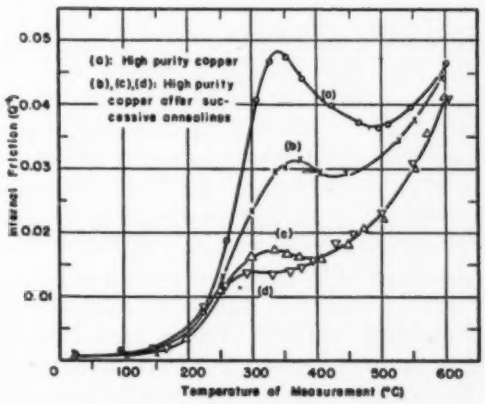


FIG. 5. Internal friction of high purity copper before and after "loaded" with oxygen.

Although the measurements were taken in an atmosphere of argon purified from oxygen, the furnace enclosing the specimen was not absolutely airtight. Accordingly, some air might have diffused

into the specimen along the grain boundaries during the measurement, especially at higher temperatures. As it is well known that oxygen diffuses readily into copper at moderately high temperatures,[22] the disappearance of the grain boundary relaxation peak is probably caused by the presence of oxygen as impurities at the grain boundaries.

Metallographic examination on the specimens in which the grain boundary slip was partly or completely blocked (curves in Fig. 3, curve (b) in Fig. 4, and curve (d) in Fig. 5) showed that there were no perceptible inclusions or precipitates at the boundary. As the amount of impurities needed to block the grain boundary slip is very small, the impurities collected at the grain boundary must be in a highly dispersed form. A small particle can be easily locked in a "disordered group" in our proposed grain boundary model, thus rendering the rearrangement of atoms in this group difficult. This will result in a blocking of the viscous slip along grain boundaries.

In conclusion, the author wants to thank Dr. C. Zener, Dr. C. S. Barrett and Dr. C. S. Smith, all of this Institute for valuable discussions.

[22] Arthur Phillips and E. N. Skinner, Trans. A.I.M.E. 143, 301 (1941).

Disk-Loaded Wave Guides

E. L. Chu and W. W. Hansen
Microwave Laboratory, Stanford University, California

(Received September 21, 1948)

Dimensions of disk-loaded wave guide suitable for use in a linear electron accelerator are calculated with high accuracy using Schwinger's method. A qualitative explanation of the method is given.

INTRODUCTION

IN the present paper we give briefly the results of some calculations on the resonant frequencies of cavities of the type shown in Fig. 1, these being of interest to us in connection with their possible use in linear accelerators. The procedures used are largely based on those of Schwinger, but since there is little available published material descriptive of Schwinger's method, it will perhaps be useful, in this first section, to give a summary of the leading idea.

The volume described by Fig. 1 is one of a considerable class for which, although one cannot construct by standard methods a solution of the wave equation satisfying the proper boundary conditions and valid throughout the whole region, one can divide the volume into two regions such that proper solutions can be found in both. Two such regions are labeled *I* and *II* in Fig. 1. For any given variation of E over the surface separating *I* and *II* one can find the values of B on either side of the dividing surface. These values naturally depend both on the given E and on the frequency, and if the correct E were known, a frequency could be found for which the B values on the two sides of the dividing surface were identical, in which case we would have an exact solution. Alternatively, we could start with an assumed B on the dividing surface, and find the E implied by this B in conjunction with the wave equation and the boundary conditions.

There have been three approaches to this problem. Hansen[1] assumed an electric field on the separating surface and matched the resulting magnetic fields on one line on this surface. In principle this is crude, but the resulting errors were minimized by good choices of electric field variation, this being made practical by a special analytical procedure.

Second, Hahn[2] and Slater[3] showed how to write an infinite array of equations determining the coefficients of a Fourier expansion of the fields. Approximate solutions of these equations are then found, Slater also showing how to make elegant and productive use of information as to the approximate form of the fields in finding these solutions.

Third, Schwinger has shown that if one starts, for example, with an assumed electric field, and adjusts the frequency, and with it the magnetic field, to a value such that the net influx of Poynting vector into the bounding surface is zero, then (a) the resulting frequency is stationary with respect to changes in assumed electric field and (b) for any specified mode, the sign of the error can be stated. And again, if one starts with an assumed magnetic field and adjusts frequency until the mean electric field on the two sides is the same, the resulting frequency is again stationary, but the error will now have the opposite sign. Here also,

[1] W. W. Hansen, J. App. Phys. 10, 38 (1939).
[2] W. C. Hahn, J. App. Phys. 12, 62 (1941).
[3] J. C. Slater, Technical Rpt. No. 48, Sept. 19, 1947, Research Laboratory of Electronics, M.I.T.

"mean electric field" means weighted according to the magnetic field, i.e., the integrated Poynting vector is zero.

Although good results are possible with either of the first two methods, Schwinger's procedure is superior, both in elegance, and in that it gives a limit on the error, and it is therefore used in what follows.

To understand this method, we start by recalling the variational procedure for solving the scalar wave equation $\nabla^2\psi+k^2\psi=0$ subject, for example, to the condition $\psi=0$ on the boundaries. Following the usual procedure, we take ψ_0 to be the solution for the lowest eigenvalue, k_0, vary ψ_0 by $\delta\psi$, and form the ratio $\int(\nabla\psi)^2d\tau/\int\psi^2d\tau$. Then, using Green's theorem and the wave equation

$$\frac{\int(\nabla\psi)^2d\tau}{\int\psi^2d\tau}=k_0^2\frac{\int\psi_0^2d\tau+2\int\psi_0\delta\psi d\tau+(2/k_0^2)\int\delta\psi\nabla\psi_0\cdot d\sigma+(1/k_0^2)\int(\nabla\delta\psi)^2d\tau}{\int\psi_0^2d\tau+2\int\psi_0\delta\psi d\tau+\int(\delta\psi)^2d\tau}. \tag{1}$$

If we now take $\delta\psi$ to be zero on the boundary, the surface integral vanishes, the other term linear in $\delta\psi$ pairs with a similar one in the denominator, and we see that the ratio is stationary about the value k_0^2. Moreover, it is easy to prove that $\int(\nabla\delta\psi)^2d\tau$ exceeds $\int(\delta\psi)^2d\tau$ so the ratio is an approximation to k_0^2 which is in error by excess.

But now suppose that $\delta\psi$ is not zero on the boundary. Then the surface integral is linear in $\delta\psi$, the stationary property is lost, and the ratio of the two integrals can have any value. The stationary property can, however, be restored by introducing the weaker condition that integral $\psi\nabla\psi$ be zero over the boundary, for this makes $\delta\psi\nabla\psi_0$ equal to $-\delta\psi\nabla\delta\psi$ and so makes the deviations quadratic in $\delta\psi$. On the other hand, the minimum property is lost.

But for a special class of trial functions, the sign of the error can be specified, and moreover, the same procedure, applied to a simply related problem, will give an error of the opposite sense. Thus it is possible to approach the true value of k_0 from both above and below, and to give limits between which it must lie.

To accomplish this, one requires the trial ψ to be an exact solution of ψ in τ. On the other hand the trial ψ does not, in general, satisfy the boundary conditions—if it did it would be ψ_0. Then in any given case, it is easy to determine the sense of the error, as will be demonstrated by example presently.[4] Moreover, the quantity $A=\nabla\psi$ satisfies a wave equation, and we can solve this in the same way except that the condition $\psi=0$ on the boundary turns into a condition on the derivatives of A. When this calculation is carried through, it turns out that the sign of the error is reversed.

[4] The sense of the error depends both on whether one works with ψ or $\nabla\psi$ and on the particular mode. The important point is that, for any given mode, either sign may be had. In a specific case it is easy to establish the sign.

In what follows we will assume that the trial function ψ is a solution of the wave equation. To emphasize that this is not a trivial assumption we may point out three important consequences.

First, it is not, in general, easy to write such solutions, so that this is, in practice, a very restrictive assumption. On the other hand, for boundaries like those of Fig. 1, this restriction is not impracticable, it being easy enough to find solutions of the wave equation in both regions I and II, but difficult to make the solutions match over the dividing surface, which here plays the same role as the boundary surface above.

Second, since the trial ψ satisfies the wave equation, the determination of the approximate k^2 from (1) is trivial—the problem is to make the surface integral of $\psi\nabla\psi$ vanish.

Third, since ψ satisfies the wave equation, the value of $\nabla\psi$ on the boundary is determined by the value of ψ, so that the problem is reduced from a three-dimensional one involving ψ everywhere inside the boundary, to a two-dimensional one involving ψ only on the separating surface.

We now make a concise statement of the problem in this form, doing this for the specific vector problem of interest to us rather than the scalar problem used in the qualitative discussion. Thus consider the electric and magnetic field components $E_z(z)$, $B_\phi(z)$ on the surface S. If, say, E_z is given, B_ϕ can be computed either from the Green's function for region I or from that for region II. It is

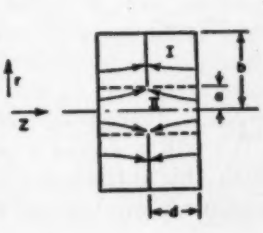


FIG. 1. Cylindrical cavity loaded by an apertured disk with electric field lines for one possible mode of oscillation. The dotted lines denote a dividing surface S which can be introduced to divide the cavity into regions I and II, in either of which solutions of the wave equation can be found.

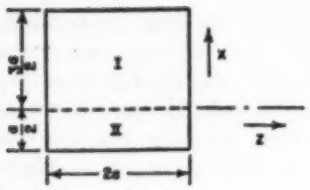


FIG. 2. End view of infinitely long square prism cavity used for simple illustration of method.

desired that the two functions B_ϕ be identical, so that, by denoting the Green's functions by G_I and G_{II}, we wish to solve the equation

$$0=\int (G_I(k;z,z')-G_{II}(k;z,z'))E_z(z')dz'$$
$$=\int G(k;z,z')E_z(z')dz'=B_{\phi I}-B_{\phi II}. \quad (2)$$

That is, G_I and G_{II} and so G being known, we search for a value of k and a function $E_z(z')$ which will make the difference of the B_ϕ, as expressed by the integral of (2), vanish for all z. Generally it will not be possible to find the proper E_z and k, but Schwinger's method asserts that if we solve

$$0=\int E\times(B_I-B_{II})dz$$
$$=\int\int E_z(z)G(k;z,z')E_z(z')dzdz' \quad (3)$$

for k, for arbitrary E_z, the resulting k^2 will approach the correct k^2 as E_z approaches the correct function, and that the deviation of k will be quadratic in the deviation of E_z. Moreover, it can be shown that if we start with B_ϕ and use a Green's function, say K, which gives E_z in terms of B_ϕ, then the same procedure will approach the true k^2 from the opposite side, and again the errors will be quadratic in the deviations of the assumed function from the true one. Thus one finds not only an approximate value for k^2, but also limits for the error.

The Eq. (3) can be given a simple and satisfying physical interpretation. For a given choice of E_z, the fields B_ϕ on the two sides of the dividing surface S are different, so that these fields can constitute only an approximate solution. But these same fields are exact solutions of a modified problem in which a current sheet of suitable strength is placed on S to support the discontinuity in B. Each filament of such a current sheet requires reactive power to maintain volt-amperes which might, for example, come from an external condenser or inductance. Equation (3) simply requires that the total reactive power required to maintain the current sheet be zero, i.e., there is no net external energy storage.

We will not give a proof of (3) but instead will finish this introductory section by illustrating by example how it comes that (3) gives too low a result, and the companion equation starting with B gives too high an answer. For this purpose we choose an example for which the correct answer is obvious, and use trial functions so specialized as to reduce analytical complexity to the vanishing point.

Thus consider the cavity of Fig. 2, this being an infinitely long square prism of sides $2a$, which we will consider as divided into regions I and II by the yz plane. The rigorous solution, for the mode to be discussed is, of course $E_z=\sin k(3a/2-x)$ in

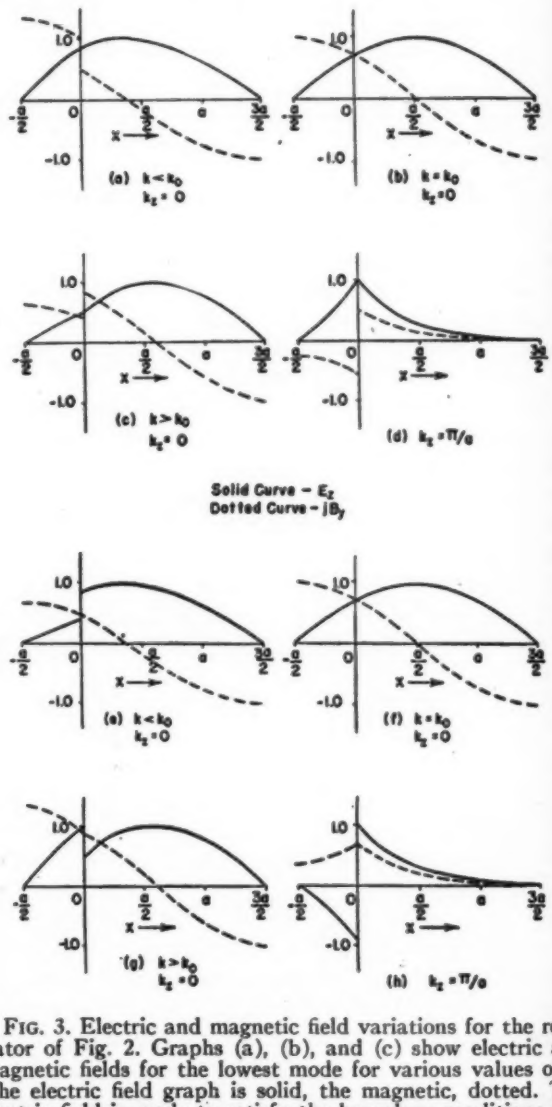


FIG. 3. Electric and magnetic field variations for the resonator of Fig. 2. Graphs (a), (b), and (c) show electric and magnetic fields for the lowest mode for various values of k. The electric field graph is solid, the magnetic, dotted. The electric field is made to satisfy the boundary conditions and to be continuous. In general, this leads to a discontinuity in the magnetic field, with sign depending on the value of k. Graph (d) shows the fields for a higher mode. Only one graph is shown since the general character is practically independent of k, i.e., the magnitude of the discontinuity varies somewhat with k, but the sign does not. Graphs (e) (f) (g) (h) are similar to those of (a) (b) (c) (d) except that boundary conditions and continuity are maintained on the magnetic rather than the electric field. Note that while the discontinuities in (d) and (h) have the same sense, the discontinuities in (a) (c) and (e) (g) are reversed.

region I, and $\sin k(a/2+x)$ in II, these joining without break on S (the yz plane) when $k=k_0=\pi/2a$ and giving $E_z=$constant on S. The third in the infinite sequence of exact solutions in I is $E_z=\cos(\pi z/a)\sinh[\{(\pi/a)^2-k^2\}^{\frac{1}{2}}(3a/2-x)]$ the solution in II being the same with $(3a/2-x)$ replaced by $(a/2+x)$. The corresponding magnetic fields come out to $j\cos k(3a/2-x)$ and $-j\{(\pi/ka)^2-1\}^{-\frac{1}{2}}\times\cos(\pi z/a)\cosh[\{(\pi/a)^2-k^2\}^{\frac{1}{2}}(3a/2-x)]$.

Let us now use some mixture of these two functions as a third trial function. Because of the orthogonality between the functions of z, it is plain that (a) if E_z, for example, is to be continuous on S, the two functions must match separately and (b) cross-product terms in $E\times B$ integrate out so that discontinuities can also be considered separately for the two functions. We therefore plot the functions separately, and show in Figs. 3a, 3b, and 3c, E_z and jB_y on any line $z=$const. for the first function and for an increasing sequence of values of k. Here the constants have been chosen to make E_z continuous at $x=0$. We observe that the sign of the discontinuity in B_y reverses as k passes through the correct value. Next, Fig. 3d, shows the fields associated with the second function. Here the curvature is always away from the axis as it will be for all higher modes and while the magnitudes vary somewhat with k, the field jB_y is always higher on the right-hand side of $x=0$. We conclude that, to get the integral of $E_z\times B_y$ to vanish, a frequency lower than the correct one must be used. If we now repeat the process but make the B_y values join, the results are as sketched in Figs. 3e, 3f, 3g, and 3h. In this case, E_z for the second function is always high on the right, as in Fig. 3h, but now the discontinuity for the first function has reversed sign, and we must go to a higher frequency. Thus we see that, as the trial E or B approaches the true one, the true value of k may be approached from either below or above and in any given case it is easy to decide from which side the approach is.

CALCULATIONS

We are interested in the mode of oscillation with field lines like those sketched in Fig. 1, and will confine ourselves to the case in which $d=\lambda/4$, $kd=\pi/2$. We note that if a number of such resonators are put end to end to make a periodic structure suitable for an electron accelerator, the closed ends may either be omitted entirely to give $\lambda/2$ loading, or replaced by apertured disks to give $\lambda/4$ loading; the frequency will be the same in either case.

Specifically, our problem is, for given ka, and $kd=\pi/2$, what value of kb will make the structure resonate?

We must first devise functions for E_z or B_ϕ that correspond as closely as possible to the actual fields, and which are analytically tractable.

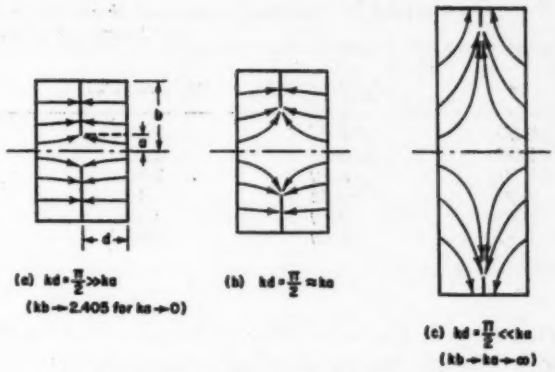


FIG. 4. Qualitative picture of the field lines for various cavity shapes. The cavities are to scale.

Information useful in this guessing process may be had by considering the limiting cases of $ka\to 0$ and $ka\to kb$. The former may be considered as two cylindrical cavities end to end, perturbed by a hole through the common end, in which case the field will be qualitatively like Fig. 4a. When $ka\to kb\to\infty$, we can consider that we have a cylindrical cavity with one axial node, and perturbed by a ring of small radial extent at $z=0$, in which case a field plot like Fig. 4c is indicated. And for intermediate values, a field like Fig. 4b would be anticipated. In all cases, E_z must approach infinity like $1/(z)^{\frac{1}{2}}$ near the edge of the disk, just as it would in a static problem, since near an edge the field curvatures imposed by satisfying the boundary conditions dominate the curvature due to the k^2 term in the wave equation, so that the wave equation is well approximated by Laplace's equation. The qualitative features of these graphs, some or all of which we hope to duplicate in the trial functions, are: (a) E_z is even about $z=d$, (b) E_z is odd about $z=0$, (c) $E_z\sim z^{-\frac{1}{2}}$ near $z=0$, and (d) for small ka E_z is nearly constant, except near $z=0$; for large ka E_z is nearly sinusoidal except near $z=0$.

Guided by the above, we choose for one function, $E_z=1$, $0<z<d$ (and the negative of this for $-d<z<0$), this being good in the limit $ka\to 0$. Next, we have used $E_z=(2dz-z^2)^{-\frac{1}{2}}$, $0<z<d$, this having the right symmetry and the right form of singularity. This function may be expected to be good for intermediate values of ka, where the field is largely controlled by the disk edge. For large ka, the function $\sin kz$ would no doubt be good, but we have not had occasion to try this, since work with a similar function for B shows that the values of ka for which this is a good choice are considerably above the range of practical interest. We have, however, felt the need for a function intermediate between the first two and for this purpose use the one parameter function $E_z=\cosh\alpha(d-z)$ with α

TABLE I. Calculated values of kb are given for various ka, the various columns corresponding to trial functions as specified at the column heads.

ka	$B_\phi=(2dz-z^2)^{\frac{1}{2}}$	$B_\phi=\sin(\pi z/2d)$	$E_z=$const.	$E_z=\cosh\alpha z$	$E_z=(2dz-z^2)^{-\frac{1}{2}}$	Uncertainty
0.50	*2.42*	2.43	2.42	*2.42*	2.29	0.00
1.00	*2.56*	2.61	2.53	*2.55*	2.46	0.01
1.50	*2.80*	2.86	2.73	*2.78*	2.75	0.02
2.00	*3.12*	3.19	3.02	3.09	*3.10*	0.02
2.50	*3.50*	3.57	3.36	3.47	*3.49*	0.01
3.00	*3.90*	3.98	3.75	3.86	*3.90*	0.00

varied to give the best value of frequency. In addition to being reasonable approximations to the expected fields, the above are also practical from the point of view of the analysis.

For approximations to the magnetic fields, similar arguments lead us to try $B_\phi=(2dz-z^2)^{\frac{1}{2}}$ for intermediate ka, and $B_\phi=\sin(\pi z/2d)$ for large ka. For small ka, the obviously suggested $B_\phi=$const. does not work, the discontinuity in current at $z=0$ implied by this choice causing all the series to diverge. But although a suitable function can be devised, it turns out not to be needed, for on trial it is found that the first function works very well indeed for small ka. Moreover, even for ka as large as 3, the square root function is better than the sine, so that in practice, only one function is needed for B_ϕ.

The actual calculations will not be recorded in detail for though complicated and tedious they are reasonably straightforward. In summary, the needed Green's functions are found in series, the Green's function for region I is combined with the Fourier series for the trial E_z (say) to give a series for B_ϕ, and this in turn is multiplied by the E_z series and integrated, and the result is compared with a similar expression obtained for region II. The resulting series involved Bessel functions in a fairly complicated way, and converge rather slowly. Numerical calculation is made possible by adding and subtracting summable series with asymptotic behavior similar to the desired series. The solutions of Maxwell's equations and the Green's functions used may be found in any of the references cited, while the method of treating the series is discussed in reference (1) and a similar spirit and technique pervades reference (3).

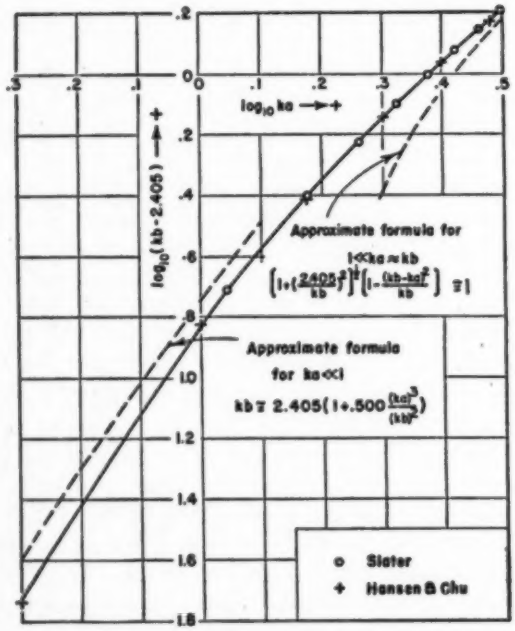


FIG. 5. Plot of $\log_{10}(kb-2.405)$ *vs.* $\log_{10}ka$. Present calculations are shown as crosses, Slater's results as circles. Also shown are two approximate formulae from an earlier paper by Chu and Hansen.

Coming now to the results, these are exhibited in Table I and Figs. 5 and 6. In the body of Table I are given approximate values of kb for various ka, the various columns corresponding to trial functions as specified at the column heads. For each value of ka the highest lower bond and the lowest upper bond are printed in italic—the true value surely lying between these limits. An upper limit for the uncertainty in kb is given in the last column.

Comparing the various trial functions, we note the following. Values of ka used do not run high enough for the sine approximation for B_ϕ to better the root function, and the discrepancy between the two is never larger than about 2 percent. The $\cosh\alpha z$ function is always better than the constant, as it obviously must be, since the former function includes the latter. But $E_z=$const. is quite good for small ka. The root function is better for large ka and might in turn be supplanted by a sine function for still larger ka. But at $ka=3.0$, no improvement is needed—the upper and lower bounds are coincident.

As to the uncertainty in ka, this is very small even if we simply take the numbers as given. And a still lower estimate is probably safe, if we note that near $ka=2.0$, where the uncertainty is a maximum, the curves for the two choices of E_z are crossing, so that the estimate from E_z is certainly too low, while on the other hand the same B_ϕ function works on both sides of this point, so the estimate from B_ϕ is probably close. We therefore believe that the

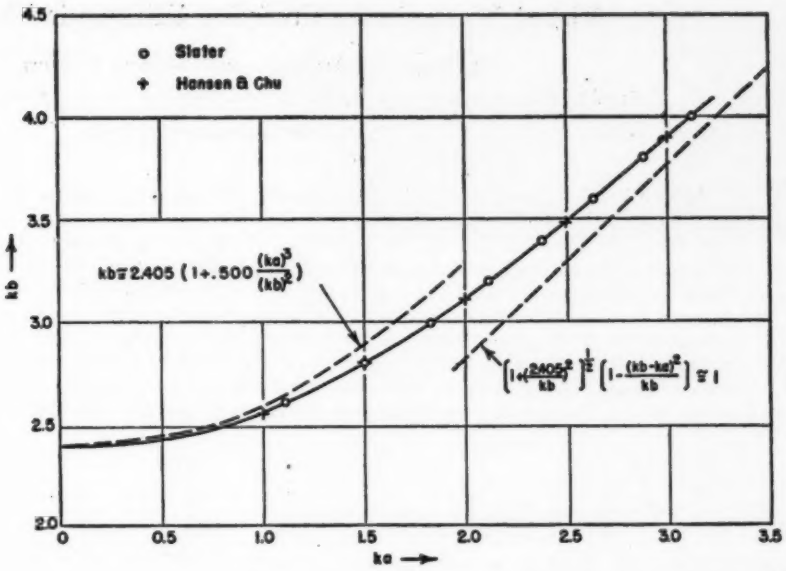


FIG. 6. Data of Fig. 5, but using linear scales.

values in the first column are everywhere the best, and that the uncertainty is about 0.01 or less.

The elegant work of Slater, mentioned in reference (3), covers exactly the same problem and gives numerical results over essentially the same range, so that a comparison is both desirable and possible. We have therefore plotted Slater's results with ours on Fig. 6, direct comparison being impossible because Slater worked with even values of kb, while we used even values of ka. It will be seen that, to within the accuracy of either the comparison or of either calculation, the two sets of results are identical. A further and nearly direct comparison is possible at one point, where Slater finds $ka=1.50$ for $kb=2.8$, while we find, for $ka=1.5$, $2.78<kb<2.80$ with the upper value preferred. Again the agreement is perfect.

CONCLUSION

Dimensions of disk-loaded wave guides suitable for use in linear electron accelerators, have been computed with high accuracy using Schwinger's method. The results are accurate to $\frac{1}{3}$ percent or better and include limits of error. Comparison with Slater's work shows perfect agreement.

Finally, we wish to acknowledge with thanks the support of the Office of Naval Research.

Visual Methods for Studying Ultrasonic Phenomena

R. Bowling Barnes* and Charles J. Burton*
Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

(Received September 20, 1948)

After a brief review of various applications of ultrasonic techniques, considerable attention is paid to the established methods employed for making sound waves visible. It is shown that these methods are primarily modifications of the technique of "schlieren photography" originally developed by Foucault and used frequently for photographing phenomena (e.g., air flow analysis) in which refractive index gradients are set up. The advantages and disadvantages of the various techniques are discussed in some detail and it is shown that a technique employed by Willard is particularly well adapted to the visualization of ultrasonic waves. A modification of this method is discussed, the experimental apparatus is described and typical photographs of ultrasonic wave patterns are shown. In particular, the close analogy existing between light waves and ultrasonic waves with respect to the wave phenomena of refraction, diffraction, and interference is demonstrated in a number of photographs.

INTRODUCTION

THE problem of detecting and measuring the wave-length, intensity, and absorption of ultrasonic vibrations has been solved in a number of ingenious ways.[1,2] Among the more important mechanical methods is the use of dust figures, the condensation of alcohol mist at the loops of stationary waves, and the measurement of sound radiation pressure by radiometers. Most important from almost every point of view, however, are the electrical and visual methods for studying these vibrations. The former utilize the piezoelectric properties of certain crystals, notably quartz or rochelle salts, or the magnetostrictive properties of various metals for generating and detecting these vibrations. The widespread use of ultrasonic transducers in sonar gear during the war is some indication of the practical application of these phenomena on a large scale.

Availing himself of Foucault's method[3] for rendering visible minute optical disturbances, Toepler[4] succeeded in observing the spherical sound waves originating in small electric sparks, and their reflection from a plane wall. Since this pioneer work, many modifications of technique have been introduced and it is a relatively simple matter to obtain visual images of sound waves in the megacycle region of the acoustic spectrum. Similarly, the remarkable developments in the field of aerodynamics, particularly since the introduction of supersonic wind tunnels, have led to far-reaching advances in the design of visual methods for air-flow analysis. It is the purpose of this paper to describe briefly some of these techniques and to show the similarities which exist in many instances between sound waves and light waves with respect to the phenomena of reflection, refraction, diffraction, and interference.

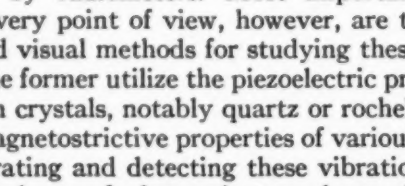


FIG. 1. Optical arrangement used by Willard for making sound waves visible.

OPTICAL METHODS OF AIR FLOW ANALYSIS

As mentioned previously, Foucault invented and Toepler perfected a very ingenious and beautiful method for making visible in a transparent medium those regions which have a refractive index differing only slightly from the surrounding regions. This method has often been referred to as the Toepler Schlieren Method. It was possible in this early work to photograph the sound waves emitted by an electric spark by employing as a light source the illumination provided by a second spark triggered an instant after the first. In the hands of Mach, Wood, and others, extensive series of photographs were prepared showing sound waves undergoing reflection, refraction and interference. An excellent review article by Barnes and Bellinger[5] describes work of this type in considerable detail.

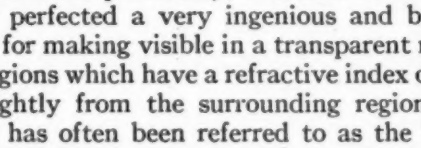


FIG. 2. Optical arrangement used in present work for making sound waves visible.

* Present address: American Optical Company, Southbridge, Massachusetts.

[1] L. Bergmann, *Ultrasonics* (John Wiley & Sons, Inc., New York 1938).

[2] E. Hiedemann, *Ultraschallforschung* (Walter De Gruyter & Company, Berlin 1939).

[3] J. L. Foucault, Ann. De L'Obs. De Paris V, 197 (1859).

[4] A. Toepler, Ann. Chim. Phys. 131, 33 (1867).

[5] N. F. Barnes and S. L. Bellinger, J. Opt. Soc. Am. 35, 497 (1945).

A number of variations of the schlieren technique have been worked out during the past few years. These later methods are primarily intended for use with ultrasonics as opposed to the schlieren methods which have been used most commonly at audible frequencies or for air flow analysis.

Primarily responsible for these developments was the almost simultaneous work of Debye and Sears[6] and Lucas and Biquard[7] who showed independently that diffraction of light may occur as a result of its passage through a liquid traversed by sound waves. The phenomenon is quite similar to that caused by a ruled optical grating and it has found frequent use as a method for the measurement of sound velocity in liquids.

Bär and Meyer[8] made use of the diffraction method to prove not only the existence of a sound wave but also to demonstrate its path and shape. Their experimental arrangement comprised substituting for the normal slit image the images of a large number of circular apertures lying in a plane. Each aperture image is formed on the final screen by light passing through an ultrasonic cell. Thus, one obtains on the viewing screen diffraction spectra on each side of the aperture image with the line joining the diffraction spectra lying in the direction of propagation of the sound wave. Consequently, one sees on the screen a diffraction image which shows the path and intensity of the sound wave.

Bachem, Hiedemann, and Asbach[9] some 15 years ago, described a method for making sound waves visible which has since become known as the "method of secondary interferences." Light coming from an illuminated slit is made parallel by a lens after which it traverses an ultrasonic cell in a direction normal to the direction of propagation of the sound wave. An objective is then focused on a plane in the sound field, and the waves become visible on the viewing screen.

Another technique of considerable importance was that developed by Hiedemann and Hoesch.[10] This comprises a variation of the schlieren method in that only a single diffracted order is allowed to pass to the final viewing screen, the direct illumination and all other diffracted orders being masked. Thus, light from an illuminated slit is made parallel by a condenser lens and is caused to pass through a sound field. Another lens focuses the light on a second movable slit which intercepts the light from one of the diffracted orders, allowing it to diverge to the viewing screen where one obtains an image of the sound field.

FIG. 3. Experimental arrangement for making visible ultrasonic beams in liquids.

Finally, one should consider briefly the method employed in the Scophony system[11] of television and by Willard[12] for making ultrasonic waves visible. As shown in Fig. 1, the light from a source of small dimensions is rendered parallel by lens L_1, passes through the ultrasonic cell and is then brought to a focus by lens L_2 at which point a stop is inserted to mask the direct illumination from the source. A third lens L_3 is placed so as to collect the light in the diffracted images and form an image of the cell on a screen S. The screen will, therefore, not be illuminated until the cell is excited, when there will appear on the screen an image of the actual sound field in the liquid. This method has been used with marked success by Willard to illustrate the effects of diffraction and interference of sound waves.

EXPERIMENTAL

Of the foregoing methods, that which seemed most suitable for our immediate purpose of making

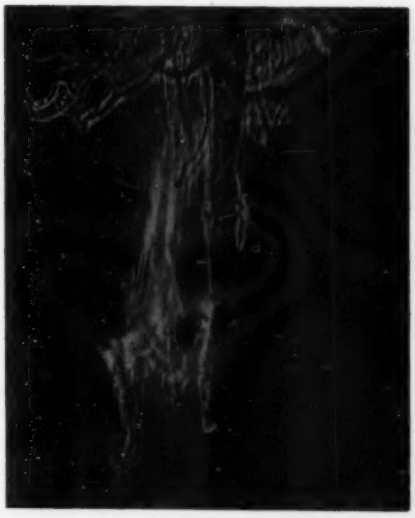


FIG. 4. Schlieren photograph of a stream of methanol squirted into water from the orifice of a micropipette immersed in the water.

[6] R. Debye and F. W. Sears, Proc. Nat. Acad. Sci., Washington 18, 410 (1932).
[7] R. Lucas and P. Biquard, Comptes Rendus 194, 2132 (1932).
[8] R. Bär and E. Meyer, Helv. Phys. Acta. 6, 242 (1933).
[9] C. Bachem, E. Hiedemann, and H. R. Asbach, Nature 133, 176 (1934).
[10] E. Hiedemann and K. H. Hoesch, Zeits. f. Physik 104, 197 (1937).
[11] F. Okolicsanyi, Wireless Eng. 14, 169 (1937).
[12] G. W. Willard, Bell Lab. Record 5, 194 (1947).

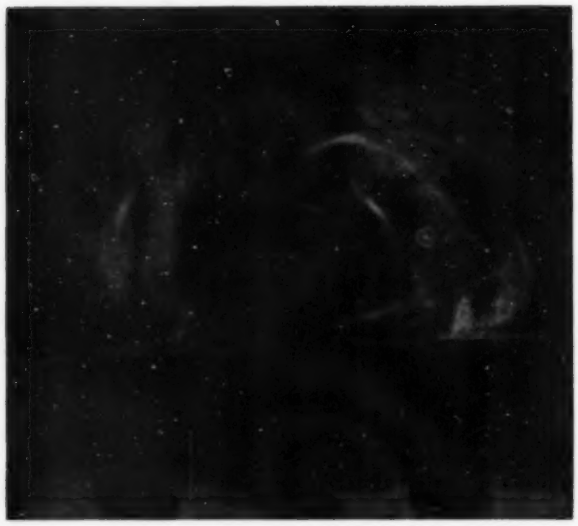


FIG. 5. Schlieren photograph of refractive index gradient at blade of immersion heater dipped in water.

visible the paths of sound waves through metal plates was a modification of the technique employed by Willard and described in the preceding paragraph.

A schematic diagram of the apparatus and a photograph of the experimental arrangement will be found in Figs. 2 and 3, respectively. It should be noted that this arrangement is identical optically with the Scophony-Willard equipment save that by judicious choice of the lenses and dimensions of the system it has been possible to eliminate the camera lens L_3 in Fig. 1.

Light from a point source of illumination (Western Union 25-watt "Point-O-Lite") is made parallel by an achromatic lens, is passed through an ultrasonic tank, and is refocused by a second lens to a point image. A stop consisting of a spot of india ink on a glass microscope slide is placed at this focal point to exclude all direct light from the camera. Some previous workers have used a slit source of illumination and a bar stop. Under these conditions, however, ultrasonic energy reflected at right angles to the direction of propagation of the sound is not visible since the vertically diffracted beams are intercepted by the bar. Use of a point source and a circular stop eliminates this difficulty and allows symmetrical observation of the various phenomena.

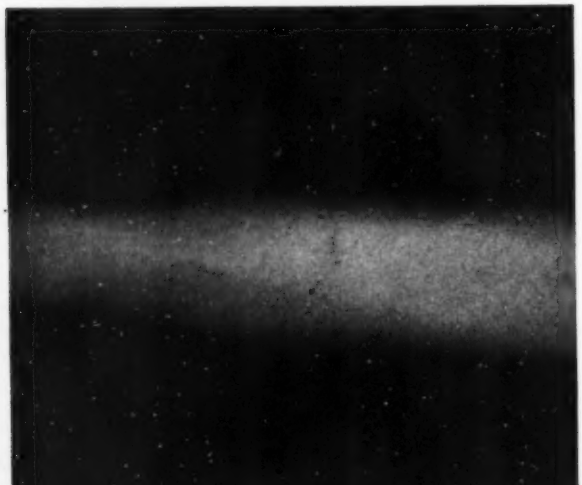


FIG. 6. A sound beam in water. Ultrasonic source is at left and a sound absorber at right.

The ultrasonic tank is fabricated from metal with plate glass windows cemented in opposite sides of the tank to allow passage of the incident light. A 5-megacycle X-cut quartz plate is mounted in a holder at one end of the tank and serves as the source of an ultrasonic beam which is caused to pass through the tank in a direction normal to that of the light beam. The ultrasonic beam is limited by a $\frac{3}{8}''$ aperture in the crystal holder and by a $\frac{1}{8}''$ slit cut in a rubber disk and supported approximately 1" in front of the holder to aid in masking the sound diffraction lobes originating at the source. The tank is considerably longer than the diameter of the lens but it can be moved with respect to the fixed optical path so that any portion of the sound field can be examined. An effort has been made to construct a tank which is reasonably anechoic. The tank has been lined with rubber and several layers of window screen have been fastened to the bottom and ends of the tank to minimize the formation of standing waves and spurious reflections. Despite these measures, some reflections from the water surface can be observed although these are not particularly troublesome.

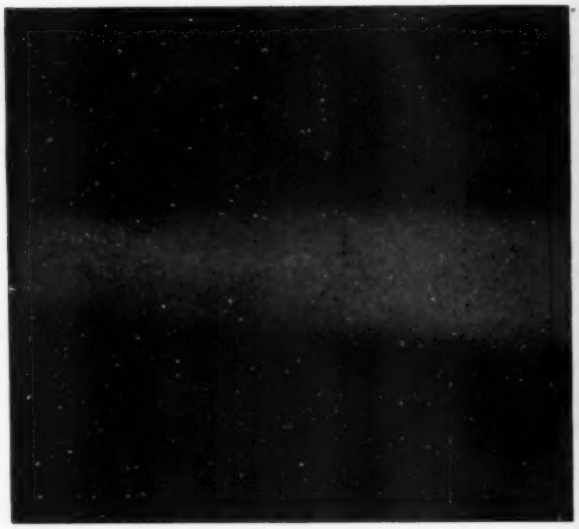


FIG. 7. A sound beam in water. Ultrasonic source is at left and a sound reflector at right.

The quartz oscillator plate is driven by an oscillator-amplifier arrangement which has been de-

scribed[13] in conjunction with some earlier ultrasonic studies. To minimize any possibility of frequency shift during operation, a second quartz crystal operating at 5 megacycles is used to control the oscillator frequency. By incorporating a link-coupled resonant circuit at the driven crystal, as much as 300 volts can be applied to the quartz, more than adequate voltage to provide a relatively intense sound beam.

Passage of the sound beam through the water produces alternate compressions and rarefactions corresponding to variations in the index of refraction of the liquid. Consequently, light passing through

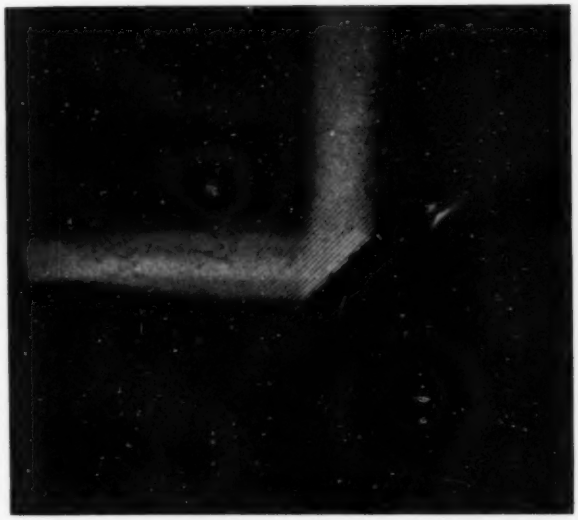


FIG. 8. Horizontal ultrasonic beam incident from left and reflected by aluminum plate oriented at 45° with respect to the incident beam.

these regions will be refracted away from the stop at the camera entrance, and an image of the actual sound beam can be projected on the camera viewing screen.

That this, strictly speaking, is merely an example of schlieren photography can be seen by an examination of Figs. 4 and 5. In Fig. 4 a stream of methanol has been squirted into the tank liquid (water) from a micropipette. Even though water and methanol are clear liquids infinitely soluble in one another, the difference in index of refraction makes the stream clearly visible. Similarly, Fig. 5 shows heat convection at the face of an immersion heater dipped in the water. The difference in refractive index between the hot and cold water creates a refractive index gradient which, as a result of the schlieren photographic technique, becomes visible on the screen. In each of these cases it should be obvious that no sound is passing through the liquid although the method of photography with or without sound is precisely the same.

[13] C. J. Burton, J. Acous. Soc. Am. 20, 186 (1948).

FIG. 9. Reflection and partial transmission of horizontal ultrasonic beam incident obliquely on an aluminum plate. Multiple transmission and reflection as a result of partial reflection is obvious.

A number of examples of ultrasonic reflection, refraction and diffraction will be found in the following section. The photographic magnification of the experimental set-up is 5¼×, and the wavelength of the sound at 5 megacycles in water is 0.3 mm. Figures 6 and 7 show the appearance of the sound beam in water with and without an absorber at the end of the tank opposite the crystal source. The standing wave pattern (Fig. 7) resulting from interference between the direct and reflected beams is obvious. At its narrowest portion the intense region of the sound beam is approximately 17–20 wave-lengths in width. Use of such a narrow sound beam produces rather pronounced sound diffraction

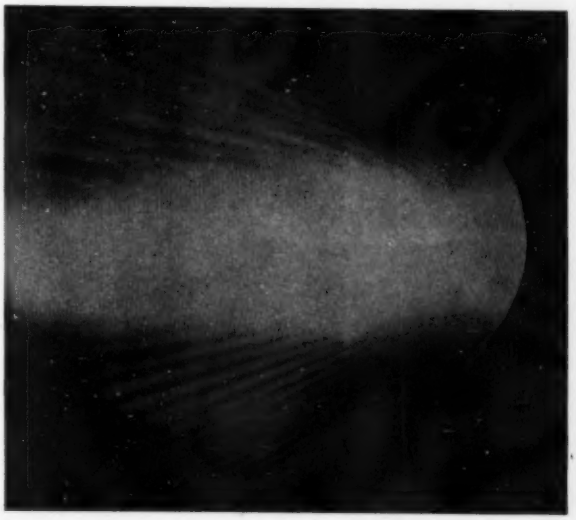


FIG. 10. Narrow ultrasonic beam incident upon a concave cylindrical reflector. Note the relatively sharp focal spot directly in front of the mirror.

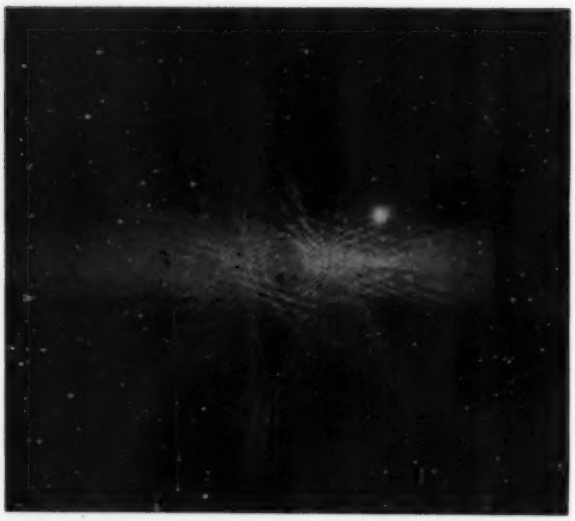


FIG. 11. Strong reflection and partial transmission of a narrow ultrasonic beam incident upon a concave cylindrical surface. The narrow entrance aperture causes much internal reflection resulting in a rather complex internal sound field. Note: the black spot on the upper lip of the reflector is an air bubble on the surface of the metal.

effects as a result of the small limiting aperture in the crystal holder. In this connection notice the beautiful diffraction effects in reference 2, Fig. 130, p. 10. Fortunately, however, these are confined fairly close to the source, and by suitable baffling of these side lobes, a relatively uniform beam of ultrasonic energy can be obtained. Diffraction effects of a similar type create serious problems in optics, particularly in astronomical studies, although the effects are of much smaller magnitude because of the relatively large dimensions of the apparatus and optical apertures when compared with the wavelength of visible light.

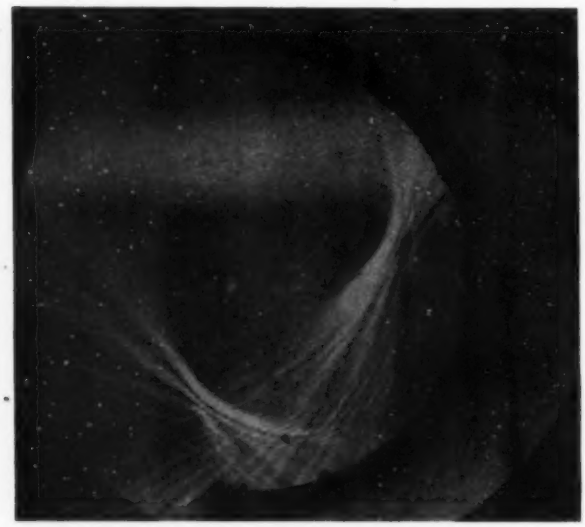


FIG. 12. Multiple internal reflection of an ultrasonic beam by a concave cylindrical surface. Note a portion of the caustic formed by reflection of the beam at a cylindrical surface.

The specular reflection of a sound beam incident from the left on a thin aluminum plate oriented at 45° with respect to the incident beam is shown in Fig. 8. Parallel to the plate one can see the fringes produced by interference of the direct and reflected beams.

A similar photograph, Fig. 9, shows not only reflection of the incident ultrasonic beam but also partial transmission of the beam with some vertical displacement as a result of refraction in the aluminum plate. Multiple transmission and reflection as a result of partial transmission and reflection within the plate is obvious.

The similarity between sound and light as wave phenomena are strikingly illustrated in Fig. 10. Here a narrow ultrasonic beam is caused to impinge upon a concave, cylindrical reflector. The focal spot

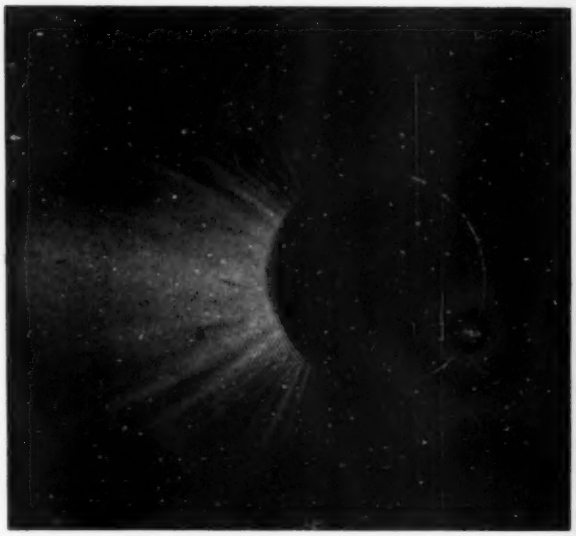


FIG. 13. Reflection of a narrow ultrasonic beam at a convex cylindrical surface. The divergent nature of the reflected beam is readily apparent.

is clearly defined as is a heterogeneous sound field directed back toward the source. The multiple reflected beams are caused by the curvature of the reflector with resultant destructive and constructive interference in the wave pattern because of path length differences. As in the optical case, the focal point of the mirror is located at one-half the radius of curvature. A similar photograph in Fig. 11 shows the complex nature of the sound field within a concave cylindrical surface with a narrow entrance aperture. A distinct focus can be seen as well as a multiplicity of reflections. Cavities of this sort have frequently been employed as radiation absorbers, the theory of operation being that much of the energy of the beam is dissipated as a result of multiple internal reflection. The photograph of the ultrasonic field supplies ample visual verification of this phenomenon.

Another type of multiple reflection is portrayed in Fig. 12 in which an ultrasonic beam is reflected considerably off-axis by a concave cylindrical reflector. One notes a portion of a typical caustic curve formed when radiation is incident upon a concave spherical or cylindrical surface.

Reflection of the ultrasonic beam at a convex cylindrical surface is shown in Fig. 13. The divergent character of the sound beam can be clearly seen. A virtual focus is obtained within the cylinder. The standing wave pattern set up between the reflector and the source is particularly well defined in this photograph as is the curvature of the wave fronts immediately adjacent to the reflecting surface.

Figures 14 and 15 are examples of multiple reflection at metallic surfaces. The interference effects in regions where waves are crossing at right angles

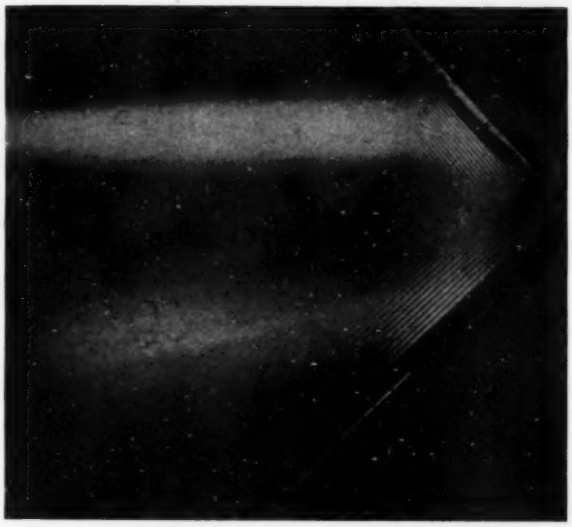


FIG. 14. Ultrasonic beam incident from left and reflected twice by right-angled aluminum reflector.

are particularly striking. The divergence of the beam after several reflections can also be seen, a result, undoubtedly, of the slight divergence of the incident beam and the comparative roughness of the reflecting surfaces.

Diffraction of light by a straight edge is one of the simplest and most striking experiments which can be performed in optics since the only apparatus required is a small source of light, a diffracting obstacle, and a screen for observation. The origin of the bands as a result of interference between the plane incident wave and the cylindrical wave initiated at the edge of the obstacle becomes obvious from a study of the ultrasonic analog shown in Fig. 16.

Figures 17 to 20 show the diffraction of sound waves by wires of varying diameters. A right-angle bend was made in a piece of wire which was im-

FIG. 15. Ultrasonic beam three times reflected at faces of rectangular cavity in brass block. Note the interference effects within the cavity as a result of interference between waves crossing at right angles.

mersed in the tank in such a manner that the portion of wire to be used as the diffracting obstacle was in the direction of the optic axis and perpendicular to the direction of propagation of the sound waves. The parabolic diffraction patterns are well defined. In Fig. 17 marked diffraction effects result from passage of the sound around a wire which is only $1/3\lambda$ in diameter.

A combination of specular reflection and diffraction at the edge of a convex cylindrical obstacle can

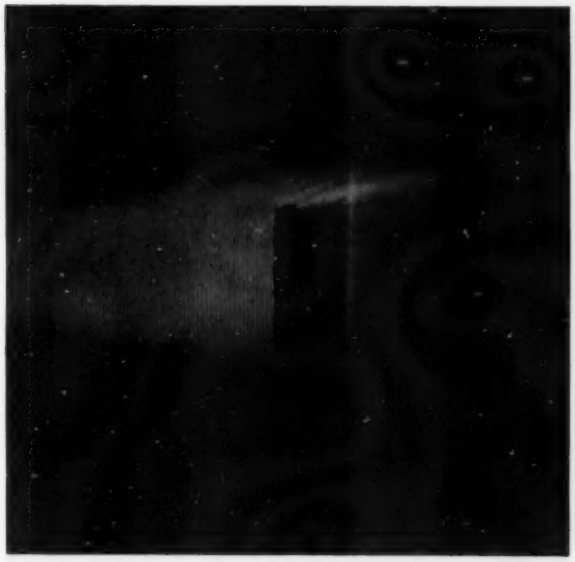


FIG. 16. Diffraction of an ultrasonic beam by a straight edge. The standing wave pattern in front of the reflector is well defined as are the parabolic diffraction patterns, these parabolae being formed by interference of a plane and a cylindrical wave front.

FIG. 17. Diffraction of an ultrasonic beam by a wire 0.11 mm (ca. $\frac{1}{3}\lambda$) in diameter. The diffracting portion of the wire is at right angles to the plane of the photograph.

be seen in Fig. 21. The curvature of the wave fronts in the proximity of the reflector is well defined. It is interesting to note the parabolic nature of these wave fronts at a considerable distance in front of the obstacle.

The experiment known as Lloyd's mirror is important in any discussion of optics and an ultrasonic analogy of this familiar experiment is shown in Figs. 22 and 23. Sound from the left is incident at a small angle upon the face of a brass reflector. Interference bands essentially parallel to the face of the reflector are produced as a result of interference between the direct and reflected beams.

FIG. 18. Diffraction of an ultrasonic beam by a wire 0.25 mm (ca. $\frac{1}{2}\lambda$) in diameter. The diffracting portion of the wire is at right angles to the plane of the photograph.

Another experiment illustrating the interference between two beams of light is known as the Fresnel mirror experiment in optics. Light from a narrow slit is split into two beams by reflection from two mirrors placed close together so that their planes make a small angle with one another. In the region where the two beams overlap, interference fringes are obtained. This phenomenon is illustrated by means of ultrasonics in Fig. 24 in which a sound beam is incident upon two thin sheets of aluminum

FIG. 19. Diffraction of an ultrasonic beam by a wire 0.44 mm (ca. $1\frac{1}{3}\lambda$) in diameter. The diffracting portion of the wire is at right angles to the plane of the photograph.

FIG. 20. Diffraction of an ultrasonic beam by a wire 0.72 mm (ca. $2\frac{1}{3}\lambda$) in diameter. The diffracting portion of the wire is at right angles to the plane of the photograph.

placed at a slight angle with one another. The interference bands in the reflected beams are clearly visible. A small amount of diffraction as a result of transmission of energy through the narrow slit between the mirrors is also evident.

One of the most fundamental optical phenomena is the diffraction of light by single and multiple slit systems. Consequently, acoustic analogies of this effect should be of particular interest. Figures 25 to 27 show diffraction of 5-megacycle ultrasonic waves by single, double, and triple slits, the slits being formed in acoustically opaque material. When immersed in water, this material buckles although the slit is still well defined. Comparison of these photographs with similar optical photographs is instructive. It can be seen that the fundamental single slit pattern is present in each of these photographs. However, as the number of slits increases, there is a narrowing of the interference maxima. In optics this effect can be carried much further, and in the case of the diffraction grating the sharpness of the principal maxima has increased to such an extent that we see "spectral lines." This effect can be shown acoustically provided a wider sound field is employed. In the present study, however, no attempt was made to go beyond a three-slit system.

FIG. 21. A combination of specular reflection and diffraction at the edge of a convex cylindrical obstacle.

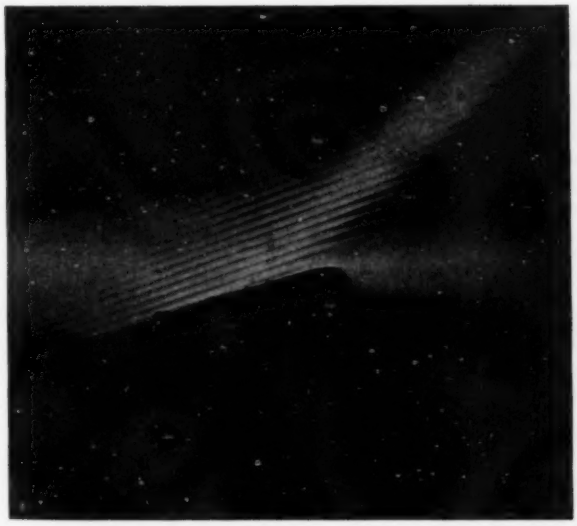


FIG. 22. Lloyd's mirror experiment showing the production of fringes as a result of interference between the direct and reflected ultrasonic beam.

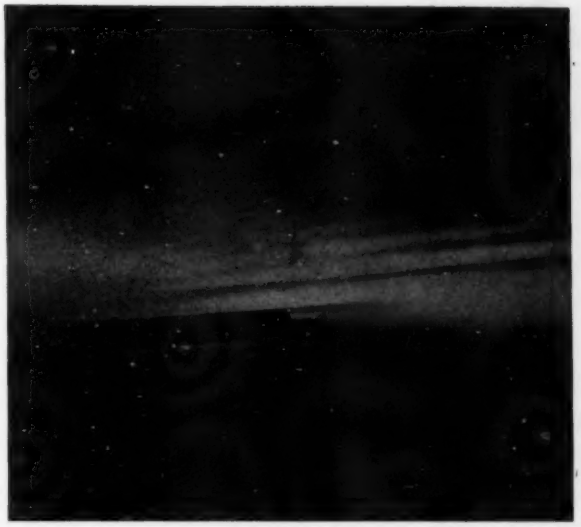


FIG. 23. Lloyd's interference fringes similar to those shown in Fig. 22, save that angle of incidence is somewhat smaller with a consequent increase in spacing of the fringes.

From the foregoing discussion it should be obvious that this optical technique for portraying the path of acoustic waves is an extremely powerful method for illustrating the mechanism of many familiar optical phenomena. The wave-length of sound in liquids is some 500–1000 times greater than that of visible light and consequently the diffraction

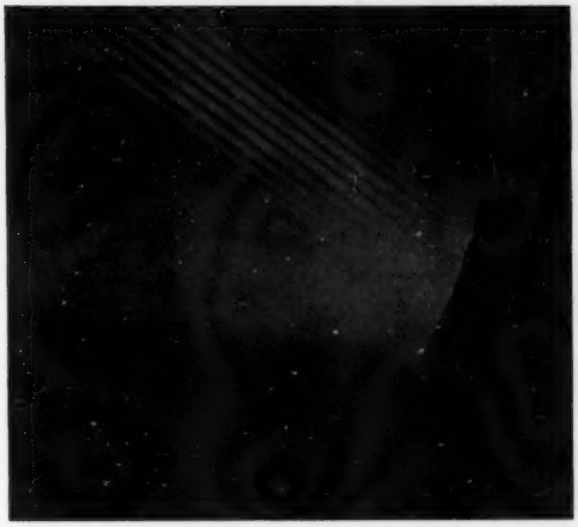


FIG. 24. Fresnel mirror experiment showing fringes produced by interference of two ultrasonic beams reflected by aluminum plates at a small angle with one another. Note also the diffraction as a result of transmission of energy through the narrow slit between the mirrors.

FIG. 25. Diffraction of a plane ultrasonic wave passing through a single slit.

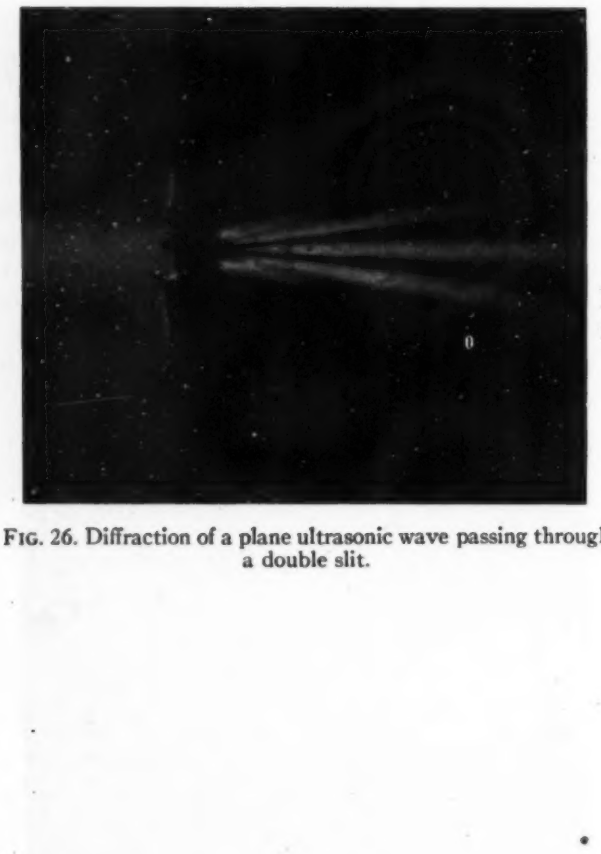


FIG. 26. Diffraction of a plane ultrasonic wave passing through a double slit.

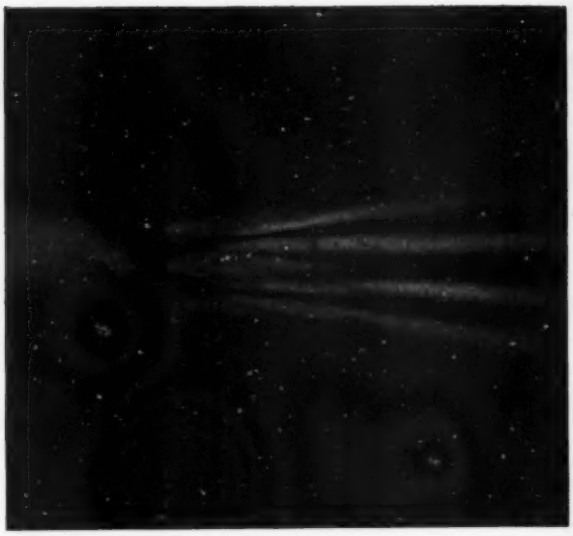


FIG. 27. Diffraction of a plane ultrasonic wave passing through a triple slit.

effects are magnified by this amount. By operating at odd harmonics of the fundamental frequency of the quartz transducer even higher frequency ultrasonic vibrations may be produced with a simultaneous decrease in the acoustic wave-length and a magnification of the diffraction phenomena. The maximum frequency of operation is limited, however, in two important respects: (1) by the low efficiency of conversion of electrical to mechanical energy by quartz oscillators at high frequencies; and (2) by the fact that the acoustic attenuation of liquids increases at high frequencies, approximately as the square of the frequency. Nevertheless, it should be possible to obtain satisfactory results at frequencies as high as 30–40 megacycles without extensive experimental difficulties. The development of a system of "ultrasonic optics" may aid considerably in illustrating many familiar optical phenomena.

Letters to the Editor

The Determination of Relative Phases of Fourier Coefficients from X-Ray Diffraction Data

DAN MCLACHLAN, JR.
Department of Metallurgy, University of Utah, Salt Lake City, Utah
October 7, 1948

THE use of x-ray diffraction as a means of determining the structure of molecules in the crystalline state using the well-known equation,

$$\rho_{xyz}=\Sigma_h\,\Sigma_k\,\Sigma_l|F_{hkl}|\cos[2\pi(hx+ky+lz)-\alpha_{hkl}],$$

has been considered a very fruitful line of approach ever since Bragg's first practical demonstration[1] of the possibilities. However, the rate at which structures have been so determined has been relatively slow because of the difficulties in determining the values of α_{hkl} or *phases* of the F values. In the case of centrosymmetric structures, α_{hkl} can take only two distinct values, 0° and 180°, which correspond to $+|F_{hkl}|$ and $-|F_{hkl}|$, respectively; therefore, for these cases reference is often made to the *signs* of the F values. Means for learning a great deal about structure without determining the phases (or signs) were developed by Patterson[2] and extended by Harker.[3] A very marked renewal of effort on the original problem, however, was recently stimulated by the work of Harker and Kasper[4,5] who have shown that by the use of Schwarz and Cauchy's inequalities, some of the signs may be determined and limits established for the phases. Buerger[6] and Gillis[7] have made contributions along these lines.

A method of treatment of x-ray data for the determination of signs or phases is herein outlined which (although making use of inequalities) differs from the other methods in that, instead of taking powers of the F values, powers are taken of the constituent parts of the expressions for the F's. This leads to new relationships between the phases or signs, and may be especially useful when used in conjunction with other methods.

Starting out with the fact that any number (a) such that $(-1)<(a)<(+1)$ when raised to an even integral power (m) gives a number a^m which is positive, and the greater the value of the power the smaller the resulting number becomes, one may conclude that a^m is positive and greater than a^{m+2}. Also a series Σa_j of such numbers may be raised term by term to even integral powers (m) and $(m+2)$ obeying the following relationship:

$$(\Sigma a_j{}^m/\Sigma a_j{}^{m+2})>1. \tag{1a}$$

And, if each term in the series Σa_j is multiplied by any characteristic positive number k_j, the resulting series $\Sigma k_j a_j$ may have its a_j terms raised to the powers (m) resulting in the series $\Sigma k_j a_j{}^m$ which is positive and decreasing in magnitude with increasing values of (m), that is,

$$(\Sigma k_j a_j{}^m/\Sigma k_j a_j{}^{m+2})>1. \tag{1b}$$

Similarly, one may start with the fact that any number (a) such that $(-1)<(a)<(+1)$ when raised to an odd power (m') gives a number $a^{m'}$ which is of the same sign as (a) (but not necessarily smaller), one may deduce that

$$(\Sigma k_j a_j{}^{m'}/\Sigma k_j a_j{}^{m+2})>0. \tag{2}$$

Since structure factors are expressed as sums of cosine and sine terms (each term being necessarily less than unity in magnitude) multiplied by an appropriate positive constant, it is convenient to have on hand the following four relationships:

$$\cos^m x=[1/2^{(m-1)}]\sum_{s=0}^{s=m/2}[C_s{}^m\cos(m-2s)x]$$
$$-[m!/2(m/2!)^2]; \tag{3}$$

$$\cos^{m'}x=[1/2^{(m'-1)}]\sum_{s=0}^{s=(m'/2-1)}C_s{}^m\cos(m'-2s)x; \tag{4}$$

$$\sin^m x=[(-1)^{(s+1)}/2^{(m-1)}]\sum_{s=0}^{s=m/2}\{C_s{}^m\cos(m-2s)x$$
$$-[m!/2(m/2!)^2]\}; \tag{5}$$

$$\sin^{m'}x=[(-1)^{(s+1)}/2(m'-1)]\sum_{s=0}^{s=(m'/2-1)}C_s{}^{m'}\sin(m'-2s)x, \tag{6}$$

where $C_s{}^m=[m!/s!(m-s)!]$.

Adopting Harker and Kasper's[4,5] definition of ${}^A\hat{F}_{hkl}$ we may illustrate the use of this method on the simplest type of problem, i.e., the relationship between F_{h00} and F_{2h00} signs in a system having symmetry along x.

$${}^A\hat{F}_{h00}=\Sigma_j\,n_j\cos 2\pi hx_j. \tag{7}$$

When $k_j=n_j$, $m'=1$, and $\cos 2\pi x_j=a_j$ are applied to Eq. (2), we find that a function that must have the same sign as ${}^A\hat{F}_{h00}$ is

$$\Sigma_j\,n_j\cos^2 2\pi hx_j, \tag{8}$$

but according to (4)

$$n_j\cos^2 2\pi hx_j=(n_j/4)(\cos 2\pi 3hx_j+3\cos 2\pi hx_j), \tag{9}$$

but by definition

$$\Sigma n_{j/4}(\cos 2\pi 3hx_j+3\cos 2\pi hx_j)=\tfrac{1}{4}({}^A\hat{F}_{3h00}+3{}^A\hat{F}_{h00}), \tag{10}$$

from which we conclude

$$(4{}^A\hat{F}_{h00}/{}^A\hat{F}_{3h00}+3{}^A\hat{F}_{h00})>0, \tag{11}$$

which means that when $|{}^A\hat{F}_{3h00}|$ is greater than $3|{}^A\hat{F}_{h00}|$, then ${}^A\hat{F}_{3h00}$ and ${}^A\hat{F}_{h00}$ must be of same sign.

The two-dimensional analog of Eq. (11) is

$$(16{}^A\hat{F}_{hk0}/{}^A\hat{F}_{3h3k0}+3{}^A\hat{F}_{3h\bar{k}0}+3{}^A\hat{F}_{h3k0})+9{}^A\hat{F}_{hk0}>0 \tag{12}$$

and reduces to (11) when $k=0$. Similarly, the more extended three-dimensional analog reduces to (12) or (11) upon appropriate assignment of zero to l or k and l. In like manner the inequalities involving the 5th, 7th, and 9th powers are found to be most useful whereas the even powers have not been very useful thus far.

The study of non-symmetrical structures is equally interesting and their discussion will be included in a more complete publication.

[1] W. L. Bragg, Proc. Roy. Soc. (London) **A123**, 537–559 (1929).
[2] A. L. Patterson, Zeits. f. Krist. **90**, 517–542 (1935).
[3] D. Harker, J. Chem. Phys. **6**, 381–384 (1936).
[4] D. Harker and J. S. Kasper, J. Chem. Phys. **15**, 882–884 (1947).
[5] D. Harker and J. S. Kasper, Acta Crystallographica **1**, 70–75 (1948).
[6] M. J. Buerger, Proc. Nat. Acad. Sci. **34**, 279–285 (1948).
[7] J. Gillis, Acta Crystallographica **1**, 76–80 (1948).

The Use of Titanium and Zirconium Nitrides as a Conducting Element in Thoria Cathodes

O. B. BUSH, R. B. VANDEGRIFT, AND T. E. HANLEY
Naval Research Laboratory, Washington, D. C.
December 17, 1948

THE ceramic thoria cathode has received much attention in the last few years because of its ability to supply high currents and to operate in high voltage tubes without arcing. However, in order to obtain sufficient primary emission from these cathodes, they must be operated at high temperatures

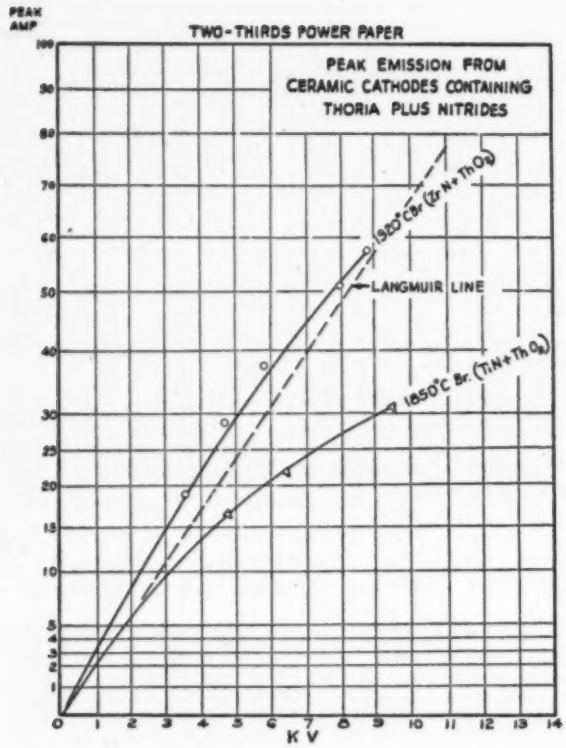


FIG. 1.

(1400–1600°C Br). Several methods of heating these ceramics have been tried, including the use of an internal tungsten spiral, heating by bombardment from an auxiliary cathode, and direct heating through the addition of a conducting material to the ceramic. Usually the conducting element used in the latter case has been a tungsten or molybdenum powder which is mixed with the thoria powder (300 mesh) in the proportion of approximately 20 percent metal powder and 80 percent thoria (by weight). The resultant combination is mixed by rolling, then pressed, and finally sintered at 1800°C Br.

At the Naval Research Laboratory a study has been made of directly heated thoria cathodes, and in the course of the investigation it has been found that addition of the nitrides of titanium and zirconium will form conducting cathodes which can be directly heated. A mixture of 20 percent of titanium nitride and 80 percent thoria has been used in rods which have been heated to a temperature as high as 2150°C Br. These rods are approximately 0.090" in diameter and have an emitting area of about 1 sq. cm. A typical rod requires 150-watt heater power to raise the temperature to 2100°C Br.

The resistance characteristics of these rods compare very favorably with the powdered metal-thoria rods. For example, one such rod containing 20 percent TiN and 80 percent ThO_2 had a resistance of 0.37 ohm at 1475°C Br and 0.48 ohm at 2030°C Br, whereas a rod containing 23 percent Mo and 77 percent ThO_2 had a resistance of 0.43 ohm at 1458°C Br and 1.07 ohms at 1932°C Br. Furthermore, these rods do not require any stabilizing or aging process to achieve a uniform resistance. The principal advantage in the use of the intrides for directly heated cathodes, however, lies in the fact that they can stand high temperatures without excessive evaporation. Rods of this type have been heated above 1900°C Br in glass envelopes without any noticeable deposits. This is a big improvement over the powdered metal cathodes, particularly the molybdenum powder thoria cathode which evaporates excessively at temperatures above 1600°C Br.

The d.c. emission of these nitride-thoria rods seems to be essentially that of pure thoria. This estimate is based on the assumption that such cathodes have a spectral emissivity of approximately 0.8, such as has been reported for powdered metal thoria cathodes. Figure 1 illustrates the peak emission obtained from thoria cathodes made conducting by the addition of zirconium and titanium nitrides. The emission shown is total emission from cathodes each having a total area of approximately one square centimeter. These tests were run with a 1-μsec. pulse and a 60-cycle rep. rate. This data is preliminary and it was obtained from diodes which did not have guard rings, therefore, comparison with other data is difficult. Furthermore, it must be emphasized that the temperatures given are brightness temperatures. No measurements of the spectral emissivities of this type of surface have yet been made.

As is quite well known, the difficult problem of ceramic cathodes is that of improving their mechanical strength. From this standpoint the TiN-ThO_2 and ZrN-ThO_2 rods are no exception. For example, one of these rods was operated at 1920°C Br, giving 40 amp./cm^2 peak current (space-charge limited) for 20 hours with no apparent drop in emission. However, the test was concluded at the end of this time because the rod opened in the middle, due to limited mechanical strength. At the present time a study is being made of the mechanical properties of these cathodes at the operating temperature.